

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 14:42:48 ; Search time 21 Seconds
(without alignments)
73.271 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79
Sequence: 1 ABGSTXDYVQNIQYAG 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	44.5	56.3	455 1	VH1HMJ nucleocapsid prote
2	43	54.4	380 2	AB6712 hypothetical prote
3	42	53.2	380 2	H95129 glucose-1-phosphat
4	42	53.2	380 2	P98000 glucose-1-phosphat
5	42	53.2	512 2	S50599 SHC1 protein - yea
6	41	51.9	169 1	B64772 yajQ protein - Esc
7	41	51.9	169 2	H50688 hypothetical prote
8	41	51.9	169 2	D85539 endo-1,4-beta-xyla
9	41	51.9	240 1	U50591 endo-1,4-beta-xyla
10	41	51.9	241 2	T7005 hypothetical prote
11	40	50.6	455 2	D81809 hypothetical prote
12	40	50.6	724 2	T23332 hypothetical prote
13	39.5	50.0	60 2	D80022 type II site-speci
14	39	49.4	313 2	A5112 metalloproteinase
15	39	49.4	576 2	A52127 protein kinase pol
16	39	49.4	623 2	G22127 B. subtilis minor
17	39	49.4	771 2	AH1209 capsid protein - h
18	39	49.4	829 2	S45048 hypothetical prote
19	38.5	48.7	545 1	T19494 amidase (EC 3.5.1.
20	38.5	48.7	796 2	J50633 vesicle protein so
21	38	48.1	149 2	S40261 hypothetical prote
22	38	48.1	229 2	D90958 hypothetical prote
23	38	48.1	250 2	B5806 glucose-1-phosphat
24	38	48.1	284 2	G71192 probable cobalt tr
25	38	48.1	323 2	G71192 nifu protein homol
26	38	48.1	497 2	B90061 hypothetical prote
27	38	48.1	526 2	AG0200 probable methyl-ac
28	38	48.1	643 2	T03518 hypothetical prote
29	38	48.1	643 2	T03518

30	38	48.1	785 2	T11719
31	38	48.1	921 2	G83902
32	38	48.1	1051 2	A38373
33	38	48.1	1146 2	S40311
34	38	48.1	1374 2	D72593
35	37.5	47.5	455 1	D45340
36	37	46.8	159 2	AB1445
37	37	46.8	175 2	S75258
38	37	46.8	210 2	C83762
39	37	46.8	240 1	S47512
40	37	46.8	312 2	D89763
41	37	46.8	315 2	A56782
42	37	46.8	327 2	T44111
43	37	46.8	372 2	AD2111
44	37	46.8	423 1	I40382
45	37	46.8	447 2	H72089

ALIGNMENTS

RESULT 1

VH1HMJ nucleocapsid protein - murine hepatitis virus (strain JHM)
C/Species: murine hepatitis virus, MHV
C/Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 24-Sep-1999
C/Accession: A04024
R/Skinner, M.A.; Siddell, S.G.
Nucleic Acids Res. 11, 5045-5054, 1983
A/Title: Coronavirus JHM: nucleotide sequence of the mRNA that encodes nucleocapsid
A/Reference number: A04024; NCID:83272950; PMID:6308569
A/Accession: A04024
A/Molecule type: genomic RNA
A/Residues: 1-455 <SKI>
A/Cross-references: GB:X00990; GB:X00757; GB:M25875; NID:G58972; PIDN:CAA25497.1; PI:
C/Comment: This genome is negative, linear, single-stranded RNA.
C/Superfamily: coronavirus nucleocapsid protein
C/Keywords: glycoprotein; nucleocapsid
F410/423/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 56.3% ; Score 44.5 ; DB 1 ; Length 455 ;
Best Local Similarity 56.2% ; Pred. No. 3.4 ;
Matches 9 ; Conservative 4 ; Mismatches 2 ; Indels 1 ; Gaps 1 ;

QY 1 ABGSTXDYVQNIQYAG 16
DB 340 ADGFTKDYVE-LQYSG 354

RESULT 2

AB6712 hypothetical protein glgC [imported] - Lactococcus lactis subsp. lactis (strain IL14
C/Species: Lactococcus lactis subsp. lactis
C/Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C/Accession: AB6712
R/Bobolotin, A.; Mincker, P.; Manger, S.; Jallion, O.; Malarme, K.; Weissenbach, J.; E
Genome Res. 11, 731-753, 2001
A/Title: The complete genome sequence of the lactic acid bacterium Lactococcus lacti
A/Reference number: AB6625; MUID:21235186; PMID:11337471
A/Accession: AB6712
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-380 <STO>
A/Cross-references: GB:AB001176; PID:G12723607; PIDN:AAK04795.1; GSPDB:GN00146
A/Experimental source: strain IL1403
C/Genetics: glgC
A/Supernfamily: glucose-1-phosphate adenylyltransferase

Query Match 54.4% ; Score 43 ; DB 2 ; Length 380 ;
Best Local Similarity 61.5% ; Pred. No. 5.3 ;
Matches 8 ; Conservative 2 ; Mismatches 3 ; Indels 0 ; Gaps 0 ;

probable vacuolar
hypothetical prote
ubiquitin-protein
integrin - fruit f
hypothetical prote
nucleocapsid prote
single-stranded DN
hypothetical prote
endo-1,4-beta-xyla
conserved hypothet
unknown protein F2
conserved hypothet
twisting motility
isocitrate dehydro
hypothetical prote

DNA Res. 8, 11-22, 2001
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gen
 A/Reference number: A9629; PMID:2116231; PMID:11258796
 A/Accession: H90688
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-169 <HAV>
 A/Cross-references: GB:BA000007; PIDN:BA833903.1; PID:g1335937; GSPDB:GN00154
 A/Experimental source: strain O157:H7, substrain RIMD 0509952
 C/Genetics:
 A/Gene: EC60480
 C/Superfamily: hypothetical protein b0426

Query Match 51.9%; Score 41; DB 2; Length 169;
 Best Local Similarity 53.3%; Pred. No. 4.8;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 EGSTXDYQNIQYAG 16
 DB 81 EGSLDYDENVTHSG 95

RESULT 8
 DB5539
 C/hypothetical protein yaJ0 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL93
 C/Species: *Escherichia coli*
 C/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C/Accession: DB5539
 R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
 Nutter 409, 529-533, 2001
 A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A/Reference number: AB5480; PMID:21074935; PMID:11206551
 A/Accession: DB5539
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-169 <STO>
 A/Cross-references: GB:AE005174; NID:g12513285; PIDN:AAG54776.1; GSPDB:GN00145; UWGP:Z05
 A/Experimental source: strain O157:H7, substrain EDL933
 C/Genetics:
 A/Gene: yaJ0
 C/Superfamily: hypothetical protein b0426

Query Match 51.9%; Score 41; DB 2; Length 169;
 Best Local Similarity 53.3%; Pred. No. 4.8;
 Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 2 EGSTXDYQNIQYAG 16
 DB 81 EGSLDYDENVTHSG 95

RESULT 9
 DB50591
 C/endo-1,4-beta-xylosylase (EC 3.2.1.8) C precursor - *Streptomyces lividans*
 N/Alternate names: xylosylase C
 C/Species: *Streptomyces lividans*
 C/Date: 14-Jul-1994 #sequence_revision 22-Nov-1996 #text_change 18-Jun-1999
 C/Accession: DB50591; PS0240
 R/Shareck, F.; Roy, C.; Yaguchi, M.; Morosoli, R.; Kluepfel, D.
 Gene 107, 75-82, 1991
 A>Title: Sequences of three genes specifying xylosylases in *Streptomyces lividans*.
 A/Reference number: US0589; PMID:92077439; PMID:1743521
 A/Accession: DB50591
 A/Molecule type: DNA
 A/Residues: 1-240 <SHA>
 A/Cross-references: GB:M64553; NID:g153530; PIDN:AAA26836.1; PID:g153531
 A/Accession: PS0240
 A/Molecule type: protein
 A/Residues: 50-80 <SH2>
 C/Genetics:
 A/Gene: xinc
 C/Function:

A/Description: catalyzes the hydrolysis of 1,4-beta-xylosidic bonds in xylans
 A/Pathway: xylan degradation
 C/Superfamily: endo-1,4-beta-xylosylase; endo-1,4-beta-xylosylase homology
 C/Keywords: extracellular protein; glycosidase; hydrolase; polysaccharide degradation
 F/1-49/Domain: signal sequence #status predicted <SIG>
 F/50-240/Product: endo-1,4-beta-xylosylase C #status experimental <MAT>
 F/62-239/Domain: endo-1,4-beta-xylosylase homology <XYL>
 F/134,226/Active site: Glu #status predicted

Query Match 51.9%; Score 41; DB 1; Length 240;
 Best Local Similarity 42.9%; Pred. No. 7.2;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14
 DB 156 SDGRTYDIQYTRY 169

RESULT 10
 DB37005
 C/endo-1,4-beta-xylosylase - *Streptomyces coelicolor*
 C/Species: *Streptomyces coelicolor*
 C/Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Jan-2000
 C/Accession: T37005
 R/Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
 submitted to the EMBL Data Library, August 1999
 A/Reference number: Z21618
 A/Accession: T37005
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-241 <OLI>
 A/Cross-references: EMBL:AL109949; PIDN:CAB52919.1; GSPDB:GN00070; SCOEDB:SCJ11.34C
 A/Experimental source: strain A3(2)
 C/Genetics:
 A/Gene: xinc; SCOEDB:SCJ11.34C
 C/Superfamily: endo-1,4-beta-xylosylase; endo-1,4-beta-xylosylase homology

Query Match 51.9%; Score 41; DB 2; Length 241;
 Best Local Similarity 42.9%; Pred. No. 7.3;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNIQY 14
 DB 157 SDGRTYDIQYTRY 170

RESULT 11
 DB1809
 C/hypothetical protein NMA1831 [imported] - *Neisseria meningitidis* (strain Z2491 serog
 C/Species: *Neisseria meningitidis*
 C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C/Accession: DB1809
 R/Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; M
 Holtroyd, S.; Jorgensen, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
 Nutter 404, 502-506, 2000
 A>Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z249
 A/Reference number: AB1775; PMID:20222556; PMID:10761919
 A/Accession: DB1809
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-455 <PAR>
 A/Cross-references: GB:AL162757; GB:AL157959; NID:g7380371; PIDN:CAB85056.1; PID:g73
 A/Experimental source: serogroup A, strain Z2491
 C/Genetics:
 A/Gene: NMA1831

Query Match 50.6%; Score 40; DB 2; Length 455;
 Best Local Similarity 61.5%; Pred. No. 23;
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 4 STXDYQNIQYAG 16
 DB 352 SADYVQRIEYAG 364

RESULT 12

hypothetical protein K04G2.3 - *Caenorhabditis elegans*
 C/Species: *Caenorhabditis elegans*
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 31-Jan-2000
 C/Accession: T23322
 R/Gardner, A.
 submitted to the EMBL Data Library, July 1996

A/Reference number: Z19727
 A/Accession: T23322
 A/Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: DNA
 A/Residues: 1-724 <WIL>
 A/Cross-references: EMBL:Z75712; PIDN:GAB00040.1; GSPDB:GN00019; CESP:K04G2.3
 A/Experimental source: clone K04G2
 C/Genetics:
 A/Genes: CESP:K04G2.3
 A/Map position: 1 347/3; 617/1; 697/1
 A/Insertions: 72/3; 1
 C/Superfamily: transitional endoplasmic reticulum ATPase, FtsH/SEC18/CDC45-type ATP-bind

Query Match
 Best Local Similarity 50.6%; Score 40; DB 2; Length 724;
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 AEGSTXDYQNT 12
 |||||
 DB 230 AEGSTSDYQL 241

RESULT 13

D98032
 type II site-specific deoxyribonuclease (EC 3.1.21.4) spnR-truncation (imported) - *Streptococcus pneumoniae*
 C/Species: *Streptococcus pneumoniae*
 C/Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
 C/Accession: D98032
 R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszczyk, L.; Burgett, S.; Dehoff, B.S.; Ee, R.; Leblanc, D.J.; Lee, L.N.; Leikowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
 Y, P.; Sun, P.M.; Winkler, M.E.
 J. Bacteriol. 183, 5709-5717, 2001
 A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
 A/Title: Genome of the Bacterium *Streptococcus pneumoniae* Strain R6.
 A/Reference number: A97872; MUID:21429245; PMID:11544234
 A/Accession: D98032
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-60 <KUR>
 A/Cross-references: GB:AE007317; PIDN:AAU0089.1; PID:G15458927; GSPDB:GN00174
 C/Genetics:
 A/Genes: spnR-truncation

Query Match
 Best Local Similarity 50.0%; Score 39.5; DB 2; Length 60;
 Matches 9; Conservative 2; Mismatches 4; Indels 5; Gaps 1;

OY 2 EGSTXDYQNT 16
 |||||
 DB 21 EGENIDLYQKGRFLYKYG 40

RESULT 14

A35122
 metalloproteinase (EC 3.4.-.-) mpr precursor, extracellular - *Bacillus subtilis*
 C/Species: *Bacillus subtilis*
 C/Date: 27-Jul-1990 #sequence_revision 27-Jul-1990 #text_change 23-Sep-2002
 C/Accession: A35122; 140010; A69660
 J. Bacteriol. 172, 1024-1029, 1990
 A/Authors: Rudolph, C.F.; Ruffo Jr., G.A.; Sullivan, B.J.; Theriault, K.A.; Ally, D.; F
 A/Title: Gene encoding a novel extracellular metalloproteinase in *Bacillus subtilis*.
 A/Reference number: A35122; MUID:90130256; PMID:2105291
 A/Accession: A35122

A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-313 <SLO>
 A/Cross-references: GB:L10505; NID:G1433209; PIDN:AAA22604.1; PID:G143210; GB:M29036
 R/Smith, H.; de Jong, A.; Bron, S.; Venema, G.
 Gene 70, 351-361, 1988

A/Title: Characterization of signal-sequence-coding regions selected from the *Bacill*
 A/Reference number: 139994; MUID:9108019; PMID:3145906

A/Accession: 140010
 A/Status: preliminary; translated from GB/EMBL/DBD
 A/Molecule type: DNA

A/Residues: 1-60,65,'L',67,'S',69,'AOA' <RES>
 A/Cross-references: GB:M22916; NID:G143701; PIDN:AAA22832.1; PID:G143702
 R/Kunt, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
 C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
 A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Evington, J.; Fabbre, C.; Ferrari,
 Nature 390, 249-256, 1997
 A/Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
 lech, J.; Harwood, C.R.; Hentat, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
 Koester, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
 A/Authors: lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
 Y, M.; Ogawa, K.; Ogilwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Scaife, Y.; Sato, T.; Scan
 A/Authors: Schleich, S.; Schuster, R.; Scottone, F.; Sekiguchi, J.; Sekowsky, A.; S
 akouchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiy
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshid
 A/Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A/Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtil*
 A/Reference number: A69580; MUID:98044033; PMID:9284377

A/Accession: A69660
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Residues: 1-313 <KUN>
 A/Cross-references: GB:299105; GB:AL009126; NID:G2632457; PIDN:CAB12018.1; PID:el182
 A/Experimental source: strain 168

C/Genetics:

A/Genes: mpr
 C/Superfamily: staphylococcal serine proteinase
 C/Keywords: hydrolase

Query Match
 Best Local Similarity 49.4%; Score 39; DB 2; Length 313;
 Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 TXDYQNTQY 14
 |||||
 DB 300 TNDVFNNTQY 309

RESULT 15

S22127
 protein kinase polo (EC 2.7.1.1.-) - fruit fly (*Drosophila melanogaster*)
 C/Species: *Drosophila melanogaster*
 C/Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1995
 C/Accession: S22127
 R/Sunkel, C.E.
 submitted to the EMBL Data Library, November 1991

A/Reference number: S22127

A/Accession: S22127

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-576 <SUN>

A/Cross-references: EMBL:X63361; NID:G8355; PIDN:CAA44963.1; PID:G8356

C/Genetics:

A/Genes: FlyBase:polo

A/Cross-references: FlyBase:FBgn0003124

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hc

C/Keywords: ATP; phosphotransferase

F/23-277/Domain: protein kinase homology <KIN>

Query Match
 Best Local Similarity 49.4%; Score 39; DB 2; Length 576;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Oy 3 GSTXDVYQNIQYA 15
| : | | | : | |
Db 551 GVSXDLVYQNIQYA 563

Search completed: January 29, 2004, 14:46:13
Job time : 23 secs

END
OF SEQUENCE INFORMATION

RESULT 9
US0591
endo-1,4-beta-xylanase (EC 3.2.1.8) C precursor - Streptomyces lividans
N/Alternate names: xylanase C
C/Species: Streptomyces lividans
C/Date: 14-Jul-1994 #sequence_revision 22-Nov-1996 #text_change 18-Jun-1999
C/Accession: US0591; PS0240
R/Shareck, F.; Roy, C.; Yaguchi, M.; Morosoli, R.; Kluepfel, D.
Gene 107, 75-82, 1991
A/Title: Sequences of three genes specifying xylanases in Streptomyces lividans.
A/Reference numbers: US0589; MUID:92077433; PMID:1743521
A/Accession: US0591
A/Molecule type: DNA
A/Residues: 1-240 <SHA>
A/Cross-references: GB:M64553; NID:G153530; PIDN:AAA26836.1; PID:G153531
A/Accession: PS0240
A/Molecule type: protein
A/Residues: 50-80 <SR2>
A/Genetics:
A/Gene: xlnC
C/Function:

C/Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C/Accession: D81809
R/Parkhill, J.; Achman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mc
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajand
Nature 404, 502-506, 2000
A/Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z249.
A/Reference numbers: A81775; MUID:20222556; PMID:10761919
A/Accession: D81809
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-455 <PAR>
A/Cross-references: GB:AL162757; GB:AL157959; NID:G7380371; PIDN:CAB85056.1; PID:G73
A/Experimental source: serogroup A, strain Z2491
C/Genetics:
A/Gene: NMA1831

Query Match	50.5%	Score 40	DB 2	Length 455
Best Local Similarity	61.5%			Pred. No. 23
Matches	8	Conservative	1	Mismatches 4; Indels 0; Gaps 0

QY 4 STDPYXNIGYAG 16
DB 352 SADDVYORIEVAG 364

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: January 29, 2004, 14:39:08 ; Search time 10 Seconds

(without alignments)
75.243 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 AEGSTXDYQNTQYAG 16

Scoring table: BLOSUM62

Searched: 127863 segs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44.5	56.3	455	1 NCAP_CVMJH	P03417 murine coro
2	43	54.4	380	1 GLGC_LACLA	O9hnl1 lactococcus
3	43	54.4	426	1 GLGC_CLOCB	O91385 clostridium
4	42	53.2	380	1 GLGC_STRPN	O97qst streptococ
5	42	53.2	387	1 GLGC_BACST	O08326 bacillus st
6	42	53.2	512	1 YB86_YEAST	P39000 saccharomyc
7	41	51.9	169	1 Y363_SHIFL	P53561 shigella fl
8	41	51.9	169	1 Y363_SHIFL	P74482 escherichia
9	41	51.9	240	1 XNXC_STRLI	P26220 streptomyce
10	39	49.4	313	1 MPR_BACSU	P39790 bacillus su
11	39	49.4	576	1 POLO_DROME	P52304 drosophila
12	39	49.4	757	1 PPR2_MOUSE	O35385 mus musculu
13	38.5	48.7	545	1 AMDS_ASPOR	O15559 aspergillus
14	38.5	48.7	796	1 VP35_HUMAN	O96qk1 homo sapien
15	38.5	48.7	795	1 VP35_MOUSE	O96qk1 mus musculu
16	38	48.1	149	1 YIIL_STRAU	P41369 straphylococ
17	38	48.1	250	1 GLGC_BACCL	P33522 bacillus ca
18	38	48.1	505	1 C4DB_DROME	O9svs79 drosophila
19	38	48.1	753	1 PPR2_HUMAN	O14830 homo sapien
20	38	48.1	785	1 VP35_SCHPO	O74552 schizosacch
21	38	48.1	1051	1 UBA1_WHEAT	P20973 triticum ae
22	38	48.1	1051	1 UBA2_WHEAT	P31251 triticum ae
23	38	48.1	1146	1 ITA1_DROME	O24247 drosophila
24	37.5	47.5	145	1 NCAP_CVMJ	P18446 murine coro
25	37	46.8	210	1 XNXC_BACST	P45705 bacillus st
26	37	46.8	383	1 CYCR_CHRVI	O82947 chromatinu
27	37	46.8	423	1 IDH_BACSU	P39126 bacillus su
28	37	46.8	604	1 VEL_HPV36	P50808 human papil
29	37	46.8	631	1 MX1_MOUSE	P09922 mus musculu
30	37	46.8	728	1 ARPA_BOOLI	P23125 escherichia
31	37	46.8	961	1 LINC_CAREL	P54336 caenorhabdi
32	36.5	46.2	454	1 NCAP_CVM3	P18447 murine coro
33	36.5	46.2	454	1 NCAP_CVM3	P03416 murine coro

34	36.5	46.2	454	1 NCAP_CVM3	P18448 murine coro
35	36	45.6	258	1 GPNA_RALSO	O87213 raietonia s
36	36	45.6	352	1 CYCR_ROSDE	P26778 roseobacter
37	36	45.6	405	1 GLCI_VIBCH	O9xids vibrio chol
38	36	45.6	772	1 GCR_RABIT	P59667 oryctolagus
39	36	45.6	1456	1 MANR_HUMAN	P22897 homo sapien
40	36	45.6	3135	1 S230_PLAFO	O08372 plasmodium
41	36	45.6	5703	1 MUSE_HUMAN	O9hc84 homo sapien
42	35.5	44.9	183	1 ITPR_MOUSE	O9d526 mus musculu
43	35	44.3	120	1 FOLB_BACSU	P28823 bacillus su
44	35	44.3	177	1 CRBS_BOVIN	P06504 bos taurus
45	35	44.3	177	1 CRBS_HUMAN	P22914 homo sapien

ALIGNMENTS

RESULT 1

NCAP_CVMJH STANDARD; PRT; 455 AA.

ID NCAP_CVMJH

AC P03417

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DB Nucleocapsid protein.

GN N.

OS Murine coronavirus (strain JHM) (MHV-JHM) (Murine hepatitis virus).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;

OC Coronaviridae; Coronavirus.

OK NCBI_TaxId=11144;

RN (1)

RP MEDLINE=83272950; PubMed=6308569;

RX Skinner M.A., Siddeall S.G.;

RT "Coronavirus JHM: nucleotide sequence of the mRNA that encodes

Nucleic Acids Res. 11:5045-5054(1983).

RL Nucleic Acids Res. 11:5045-5054(1983).

CC -1 SIMILARITY: Belongs to the coronavirus nucleocapsid protein

family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC EMBL; X00990; CAA5497.1; -.

DR PIR; A04024; VHIHMJ.

DR InterPro; IPR001218; Corona_nucleocap.

DR Pfam; PF00937; Corona_nucleoca; 1.

SW Nucleocapsid.

QY SEQUENCE 455 AA; 49714 MW; 36FC529D127B5B5B CRC64;

Db 340 ADGPTXDYVE-LQYSG 354

Query Match 56.3%; Score 44.5; DB 1; Length 455;

Best Local Similarity 56.2%; Pred. No. 1.4;

Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;

RESULT 2

GLGC_LACLA STANDARD; PRT; 380 AA.

ID GLGC_LACLA

AC Q9CHN1

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Glucose-1-phosphate adenyllyltransferase (EC 2.7.7.27) (ADP-glucose

synthase) (ADP-glucose pyrophosphorylase) (ADPglc Ppase).

GN GLGC OR L10697.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 NCBI_TaxID=1360;
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=11403;
 RX MEDLINE=2135186; PubMed=11337471;
 RA Bolotin A., Wincker P., Manger S., Jallion O., Malarme K.,
 Weisenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 lactis ssp. lactis 11403".
 RL Genome Res. 11:731-753 (2001).
 CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
 CC diphosphate + ADP-glucose.
 CC -1- PATHWAY: Glycogen biosynthesis; first step.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
 CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; A006303; AK04795.1; -
 CC DR PIR; A86712; A86712.
 CC DR HAMAP; MF_00624; -; 1.
 CC DR InterPro; IPR005836; ADP_Glu_pyroph.
 CC DR InterPro; IPR001451; Hexaprep transf.
 CC DR InterPro; IPR005835; NTP transferase.
 CC DR Pfam; PF00132; hexaprep; 3.
 CC DR Pfam; PF00483; NTP transferase; 1.
 CC DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
 CC DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; FALSE NEG.
 CC DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; 1.
 CC KW Glycogen biosynthesis; transferase; Nucleotidyltransferase;
 CC Complete proteome.
 CC SQ SEQUENCE 380 AA; 42053 MW; 524B9934E6C6F68 CRC64;
 CC -----
 CC Query Match 54.4%; Score 43; DB 1; Length 380;
 CC Best Local Similarity 61.5%; Pred. No. 2.2;
 CC Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 CC -----
 CC OY 2 EGSTXDYQNIQY 14
 CC Db 100 EGSHAVYQNIQY 112
 CC -----
 CC RESULT 3
 CC GLGC_CLOCE STANDARD; PRT; 426 AA.
 CC ID GLGC_CLOCE
 CC AC Q9L385; Rel. 42, Created)
 CC DT 15-SEP-2003 (Rel. 42, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
 CC synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Phase).
 CC GN GlcC
 CC OS Clostridium cellulolyticum.
 CC OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 CC Clostridium.
 CC OX NCBI_TaxID=1521;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=ATCC 35319;
 CC RA Ravagnani A., Jernert K.C.B., Steiner E., Gruenberg R., Jeffries J.R.,
 RA Wilkinson S.R., Young M., Tidswell E.C., Brown D.P., Youngman P.V.,
 RA Morris J.G., Young W.;
 RT "Sp0A directly controls the switch from acid to solvent production in
 RT solvent-forming clostridia.";

RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
 CC diphosphate + ADP-glucose.
 CC -1- PATHWAY: Glycogen biosynthesis; first step.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
 CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AJ277601; CAB89282.1; -
 CC DR HAMAP; MF_00624; -; 1.
 CC DR InterPro; IPR005836; ADP_Glu_pyroph.
 CC DR InterPro; IPR005835; NTP transferase.
 CC DR Pfam; PF00483; NTP transferase; 1.
 CC DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
 CC DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
 CC DR PROSITE; PS00810; ADP_GLC_PYROPHOSPH_3; FALSE NEG.
 CC KW Glycogen biosynthesis; transferase; Nucleotidyltransferase.
 CC SQ SEQUENCE 426 AA; 47094 MW; B56B7B34E137A0D CRC64;
 CC -----
 CC Query Match 54.4%; Score 43; DB 1; Length 426;
 CC Best Local Similarity 61.5%; Pred. No. 2.5;
 CC Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
 CC -----
 CC OY 2 EGSTXDYQNIQY 14
 CC Db 101 KGTANAVYQNIQY 113
 CC -----
 CC RESULT 4
 CC GLGC_STREP STANDARD; PRT; 380 AA.
 CC ID GLGC_STREP
 CC AC Q97067;
 CC DT 15-SEP-2003 (Rel. 42, Created)
 CC DT 15-SEP-2003 (Rel. 42, Last sequence update)
 CC DT 15-SEP-2003 (Rel. 42, Last annotation update)
 CC DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
 CC synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Phase).
 CC GN GlcC OR Sp112.
 CC OS Streptococcus pneumoniae.
 CC OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 CC OX NCBI_TaxID=1313;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=ATCC BAA-334 / TIGR4;
 CC RX MEDLINE=21357209; PubMed=11463916;
 CC RA Tetteelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayah L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Uitterlinden T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 RA Holt I.E., Loftus B.V., Yang F., Smith H.O., Venter J.C., C.M.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Frazer J.C.,
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae".
 RT Science 293:498-506 (2001).
 CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
 CC diphosphate + ADP-glucose.
 CC -1- PATHWAY: Glycogen biosynthesis; first step.
 CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
 CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its


```

used by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.tsb-sib.ch/announce/
or send an email to license@tsb-sib.ch).
CC -----
DR EMBL, AE007413; AAK75233.1; -.
DR PIR, H95129; H95129.
DR TIGR, SPL122; -.
DR HAMAP, MF_00624; -, 1.
DR InterPro, IPR005836; ADP_Glu_pyrop.
DR InterPro, IPR005835; NTP_transferase.
DR Pfam, PFO0483; NTP_transferase; 1.
DR PROSITE, PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE, PS00809; ADP_GLC_PYROPHOSPH_2; FALSE NEG.
DR PROSITE, PS00810; ADP_GLC_PYROPHOSPH_3; FALSE NEG.
DR GlycoGen biosynthesis; Transferase; Nucleotidyltransferase;
KM Complete proteome.
SQ SEQUENCE 380 AA; 41490 MW; 23661026D2D04309 CRC64;
QY 2 EGSTXDYQNTQY 14
Db 100 EGSTSHAYQNDY 112
RESULT 5
GLGC_BACST STANDARD; PRT; 387 AA.
ID _GLGC_BACST
AC 008326;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27) (ADP-glucose
DE synthase) (ADP-glucose pyrophosphorylase) (ADPGlc Phase).
GN GLGC.
OS Bacillus stearothermophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1422;
[1]
RN R1
RP SEQUENCE FROM N.A.
RX STRAIN=TRBE24;
RX MEDLINE=97386405; PubMed=9244254;
RX Takata H., Takaha T., Okada S., Takagi M., Imanaka T.;
RT "Characterization of a gene cluster for glycogen biosynthesis and a
RT heterothermophilic ADP-glucose pyrophosphorylase from Bacillus
RT stearothermophilus."
RL J. Bacteriol. 179:4689-4698(1997).
CC -1- CATALYTIC ACTIVITY: ATP + alpha-D-glucose 1-phosphate =
CC diphosphate + ADP-glucose.
CC -1- PATHWAY: Glycogen biosynthesis; first step.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL AND PLANTS GLUCOSE-1-
CC PHOSPHATE ADENYLYLTRANSFERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.tsb-sib.ch/announce/
CC or send an email to license@tsb-sib.ch).
CC -----
DR EMBL, D97026; EAA15689.1; -.
DR HAMAP, MF_00624; -, 1.
DR InterPro, IPR005836; ADP_Glu_pyrop.
DR InterPro, IPR005835; NTP_transferase.
DR Pfam, PFO0483; NTP_transferase; 1.
DR PROSITE, PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE, PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR PROSITE, PS00810; ADP_GLC_PYROPHOSPH_3; 1.
DR GlycoGen biosynthesis; Transferase; Nucleotidyltransferase.
KW

```

```

SQ SEQUENCE 387 AA; 43268 MW; 3660DB962442D97B CRC64;
Query Match 53.2%; Score 42; DB 1; Length 387;
Best Local Similarity 61.5%; Pred. No. 3.5;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSIXDYQNIQY 14
   ||| |||||
   ||| |||||
Db 100 EGTANVYQNIY 112

RESULT 6
ID YES6_YEAST STANDARD; PRT; 512 AA.
AC P39070;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical 56.6 kDa protein in RAD51-UBP9 intergenic region.
GN YER096W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
[1]
SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX PubMed=9169868;
DI Dietrich F.S., Mulligan J.T., Hennesey K.M., Yelton M.A., Allen E.,
RA Arunjo R., Aviles E., Berno A., Brennan T., Carpenter J., Chen E.,
RA Cherry J.M., Chung E., Duncan M., Guzman E., Hartzell G.,
RA Hunkle-Smith S., Hyman R.W., Kayser A., Komp C., Lashkari D., Lew H.,
RA Lin D., Mossdale D., Nakahara K., Nishikawa A., Nogren R., Oefner P.,
RA Oh C., Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T.,
RA Smith V., Taylor P., Wei V., Botstein D., Davis R.W.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome V.";
RL Nature 387:78-81(1997).
[2]
SEQUENCE OF 248-454 FROM N.A.
RP MEDLINE=9225787; PubMed=1581961;
RX Shinozuka A., Ogawa H., Ogawa T.;
RT "Rad51 protein involved in repair and recombination in S. cerevisiae
is a RecA-like protein.";
RL Cell 69:457-470(1992).
CC -1. SIMILARITY: SOME. TO YEAST SKTS AND TO S.POMBE SPAC24B11.10C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U18839; AAB64651.1; -
DR EMBL; D10023; BAA20966.1; ALT_SEQ.
DR PIR; S50599; S50599.
DR SGD; S0000898; SHC1.
DR GO; GO:0008047; F-enzyme activator activity; IGI.
DR GO; GO:0006037; Pichitin biosynthesis; IGI.
DR GO; GO:0030437; P:sporulation (sensu Fungi); IDA.
DR InterPro; IPR006597; Sel1 like.
DR SMART; SM00671; Sel1; 4.
DR Hypothetical protein.
KW CONFLICT 311 311 K -> R (IN REF. 2).
FT SEQUENCE 512 AA; 56578 MW; 654D742482616658 CRC64;

Query Match 53.2%; Score 42; DB 1; Length 512;
Best Local Similarity 63.6%; Pred. No. 4.7;
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 TXDYQNIQYA 15
   ||| |||||
   ||| |||||

```

Db 255 TQDIYSNFOYA 265

RESULT 7
Y363_SHIFL STANDARD; PRT; 169 AA.
ID Y363_SHIFL STANDARD; PRT; 169 AA.
AC P59561;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Hypothetical UPF0234 protein ya3Q.
GN Y363 OR SF0363.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Shigella.
ON NCBI_TaxID=623;
RX MEDLINE=2272406; PubMed=12384590;
RA Jin Q., Yuan Z., Xu J., Wang Y., Shen Y., Lu W., Wang J., Liu H.,
Yang J., Yang F., Zhang X., Zhang J., Yang G., Wu H., Qu D., Dong J.,
Sun L., Xue Y., Zhao A., Gao Y., Kan B., Ding K., Chen S.,
Chang H., Yao Z., He B., Chen R., Ma D., Qiang B., Wen Y., Hou Y.,
Yu J.;
RT "Genome sequence of Shigella flexneri 2a: insights into pathogenicity
through comparison with genomes of Escherichia coli K12 and O157.";
RL Nucleic Acids Res. 30:4432-4441(2002).
CC -1- SIMILARITY: Belongs to the UPF0234 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE015069; AAN42021.1; ALT_INIT.
DR HAMAP; MF 00632; -; 1.
KW Hypothetical protein.
SQ SEQUENCE 169 AA; 19139 MW; 36A27E9CEDBD1BDC CRC64;

Query Match 51.9%; Score 41; DB 1; Length 169;
Best Local Similarity 53.3%; Pred. No. 2.1;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Cy 2 EGSFTXDVYONIQVAG 16
Db 81 EGSSLDVPENVHSG 95

RESULT 8
Y363_SHIFL STANDARD; PRT; 169 AA.
ID Y363_SHIFL STANDARD; PRT; 169 AA.
AC P77462;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE UPF0234 protein ya3Q.
GN Y363 OR B0426 OR C0537 OR Z0529 OR EGS0480.
OS Escherichia coli.
OS Escherichia coli O6, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
ON NCBI_TaxID=562; 217992; 833334;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
MAU B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
Duncan M., Federpriel N., Hyman R., Kalman S., Komp C., Kurdi O.,
Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.,
Submitted (Jan-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O6:H1 / CFT073 / ATCC 700928;
RX MEDLINE=22388234; PubMed=12471157;
RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,
Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,
Wayne G.F., Rose D.C., Zhou S., Schwartz D.C., Perna N.T.,
Mobley H.L.T., Donnenberg M.S., Blattner F.R.;
RT "Extensive mosaic structure revealed by the complete genome sequence
of uropathogenic Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
Roodca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
Weich R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;
RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
hydroxyapatite chromatography.";
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: Belongs to the UPF0234 family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE00149; AAC73529.1; -
DR EMBL; U82664; AAB40182.1; -
DR EMBL; AE016756; AAN79015.1; ALT_INIT.
DR EMBL; AE005222; AAG54776.1; -
DR EMBL; AP002551; BAB33503.1; -
DR PIR; B64772; B64772.
DR PIR; B64772; B64772.
DR PIR; D85539; D85539.
DR PIR; H90688; H90688.
DR EcoGene; EG31613; ya3Q.
DR HAMAP; MF 00632; -; 1.
DR Pfam; PF04461; DUF520; 1.
KW Complete proteome.

SEQ SEQUENCE 169 AA; 19047 MW; 202279CC84D155C CRC64;

Query Match 51.9%; Score 41; DB 1; Length 169;

Best Local Similarity 53.3%; Pred. No. 2.1; Mismatches 8; Conservative 4; Indels 3; Gaps 0;

QY 2 EGSTXDYQNIQVYG 16
ID ID
DB 81 EGSLDVENIVHSG 95

RESULT 9

XTNC_STRLI STANDARD; PRT; 240 AA.

AC P26220;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endo-1,4-beta-xylanase C precursor (EC 3.2.1.8) (Xylanase C)
DE (1,4-beta-D-xylan xylanohydrolase C).
GN XLNC.

OS Streptomyces lividans.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomyces; Streptomyces; Streptomyces.

OX NCBI_TaxID=1916;

RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 50-80.

RC STRAIN=6 / 1326;
RX MEDLINE=92077439; PubMed=1743521;

RA Shaeck F., Roy C., Yaguchi M., Morosoli R., Kluepfel D.;
RT "Sequences of three genes specifying xylanases in Streptomyces
lividans";

RU Gene 107:75-82 (1991).
CC -1- FUNCTION: Contributes to hydrolyze hemicellulose, the major
CC component of plant cell-walls.

CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-xylosidic
CC linkages in xylans.

CC -1- PATHWAY: Xylan degradation.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY G (FAMILY 11 OF GLYCOSYL
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; M64553; AAA26836.1; -

DR EMBL; A25307; CA01768.1; -

DR PIR; J30591; J30591.

DR HSSP; P09650; 1XNB.

DR InterPro; IPR001337; Glyco_hydro_11.

DR InterPro; IPR006311; Tac.

DR Pfam; PF00457; Glyco_hydro_11; 1.

DR PRINTS; PRO0911; GLHYDRLASE11.

DR TIGRPFAMS; TIGR01409; TAR signal seq; 1.

DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.

DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.

KM Xylan degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 49

FT CHAIN 50 240

FT ACT SITE 134 134 ENDO-1,4-BETA-XYLANASE C.

FT ACT SITE 226 226 NUCLEOPHILE (BY SIMILARITY).

FT ACT SITE 226 226 PROTON DONOR (BY SIMILARITY).

SEQ SEQUENCE 240 AA; 25673 MW; FC663415780142CA CRC64;

Query Match 51.9%; Score 41; DB 1; Length 240;

Best Local Similarity 42.9%; Pred. No. 3.1; Mismatches 6; Conservative 4; Indels 4; Gaps 0;

QY 1 AEGSTXDYQNIQY 14

DB 156 SDGTYDXYQTRY 169

QY 313 AA.

ID ID

AC P39790;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Extracellular metalloprotease precursor (EC 3.4.21.-).

GN MPR.

OS Bacillus subtilis.

OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

OX NCBI_TaxID=1423;

RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC STRAIN=GP241;

RA MEDLINE=90130256; PubMed=2105291;

RA Sloma A., Rudolph C.F., Rufe G.A. Jr., Sullivan B.T., Theriault K.A.,
RA Ally D., Pero J.;

RT "Gene encoding a novel extracellular metalloprotease in Bacillus
RT subtilis";

RU J. Bacteriol. 172:1024-1029 (1990).

RN [2]
RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA Haga K., Liu H., Yasunoto K., Takahashi H., Yoshikawa H.;

RT "Sequence analysis of the 70kD region between 17 and 23 degree of the
RT Bacillus subtilis chromosome.";

RU Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.

RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=168;

RA MEDLINE=96044033; PubMed=9384377;

RA Kunz F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertorello M.G., Bessieres P., Bolotin A., Borchert S.,
RA Bories R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denton F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabre C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Gallitz A., Galleron N.,
RA Gilm S.Y., Glaser P., Goffeau A., Goldlighty E.J., Grandi G.,
RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holzapfel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerner-Bianchand M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krog S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Muel C., Medigue C.,
RA Medina N., Meliado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogasawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Portwollik S., Prescott A.M.,
RA Priescan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,
RA Sato T., Scamian E., Schleich S., Schroeder R., Scifone F.,
RA Sekiguchi J., Sekowski A., Seror S.J., Seror P., Shin B., Solde B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeda M., Tanaka T., Teruya P., Tognoni A.,
RA Toso U., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler K., Wetzengger T.,
RA Winer S., Wipat A., Yamamoto H., Yarnes K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumschein E., Yoshikawa H., Darchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";

RL Nature 390:249-256 (1997).

RP SEQUENCE OF 1-68 FROM N.A.

RA MEDLINE=89106019; PubMed=3145906;

RT Smith H., de Jong A., Bron S., Venema G.;

RT "Characterization of signal-sequence-coding regions selected from the
RT Bacillus subtilis chromosome.";

```

RL Gene 70:351-361 (1988).
RN [5]
RP SEQUENCE OF 1-10 FROM N.A.
RX STRAIN=168;
RA MEDLINE=96118702; PubMed=7496533;
RA Saxild H.H., Jacobsen J.H., Nygaard P.;
RT "Functional analysis of the Bacillus subtilis purT gene encoding
RT formate-dependent glycinamide ribonucleotide transformylase.";
RL Microbiology 141:2211-2218 (1995).
CC
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2B.
CC -1- CAUTION: CALLED "METALLOPROTEINASE", BUT CLEARLY BELONGS TO THE S2
CC FAMILY OF SERINE PROTEASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10505; AAA22604.1; -
DR EMBL; AB006424; BAA33121.1; -
DR EMBL; Z99105; CAB12018.1; -
DR EMBL; M22916; AAA22832.1; -
DR PIR; A35122; A35122.
DR MEROPS; S01.272; -
DR Subtilisin; BGI0690; mpr.
DR InterPro; IPR000126; Ser_proteas_V8.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00020; TRY_SPC; 1.
DR PROSITE; PS00672; V8_HIS; 1.
DR PROSITE; PS00673; V8_SER; 1.
DR Hydroxylase; Serine protease; Signal; Zymogen; Complete proteome.
FT SIGNAL 1 34
FT PROPEP 35 93
FT CHAIN 94 313 EXTRACELLULAR METALLOPROTEASE.
FT ACT_SITE 146 146 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 267 267 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 131 147 BY SIMILARITY.
FT CONFLICT 61 68 OVSAPYEG -> PLESTAOA (IN REF. 4).
SQ SEQUENCE 313 AA; 33842 MW; D411888ED652AE94 CRC64;
Query Match 49.4%; Score 39; DB 1; Length 313;
Best Local Similarity 70.0%; Pred. No. 9.9;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
CY 5 TXDYYNIOY 14
DB 300 TNDVFNIOY 305

```

```

RA Llamazares S., Moreira A., Tavares A., Girdham C., Spruce B.A.,
RA Gonzalez C., Karsess R.E., Glover D.M., Sunkel C.E.;
RT "Polo encodes a protein kinase homolog required for mitosis in
RT Drosophila.";
RL Genes Dev. 5:2153-2165 (1991).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA Amandas S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abilov J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Bencs P.V., Bertram B.P., Bhandari D., Bolintsov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport J.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Gliedek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.U., Wei M.-H., Itoigawa C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemnson U.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mialheina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusken D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,
RA Williams S.W., Woodase T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao C., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou M., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
CC -1- FUNCTION: MAY PLAY A ROLE IN REGULATING BOTH NUCLEAR AND
CC CYTOPLASMIC ASPECTS OF THE MITOTIC CYCLE.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic. BRAIN AND TESTIS.
CC -1- TISSUE SPECIFICITY: LARVAL DISCS.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC CDC5/POLO SUBFAMILY.
CC -1- SIMILARITY: Contains 2 POLO box domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X63361; CAA44963.1; -
DR EMBL; AE003514; AAF4936.1; -
DR PIR; S22127; S22127.
DR HSSP; Q63450; 1A06.
DR FLYBase; FBgn0003124; polo.
DR GO; GO:0005813; C:centrosome; IDA.
DR GO; GO:00005813; C:spindle; IDA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IDA.
DR GO; GO:0007140; F:male meiosis; IMP.
DR GO; GO:0007067; F:mitosis; IMP.

```

DR InterPro; IPR000959; POLO_box.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR002290; Ser_thr_kinase.
 DR Pfam; PF00069; Kinase; 1.
 DR Pfam; PF00659; POLO_box; 2.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR PROSITE; PS50078; POLO_BOX; 2.
 DR PROSITE; PS50107; PROTEIN KINASE ATP; FALSE_NEG.
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
 DR PROSITE; PS00108; PROTEIN KINASE ST; 1.
 KW Transferrase; Serine/threonine-protein kinase; ATP-binding; Repeat.
 FT DOMAIN 25 277 PROTEIN KINASE.
 FT NP_BIND 31 39 ATP (BY SIMILARITY).
 FT BINDING 54 54 BY SIMILARITY).
 FT ACT_SITE 148 148 BY SIMILARITY).
 FT DOMAIN 398 461 POLO_BOX 1.
 FT DOMAIN 496 564 POLO_BOX 2.
 FT CONFLICT 187 187 P -> A (IN REF. 1).
 SQ SEQUENCE 576 AA; 66973 MW; 5022B9AC0E888FAD CRC64;

Query Match 49.4%; Score 39; DB 1; Length 576;
 Best Local Similarity 53.8%; Pred. No. 19;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSTXDVYQNIQYV 15
 DB 551 GVSXDLVQXIRYA 563

RESULT 12
 PP2_MOUSE STANDARD; PRT; 757 AA.
 ID PP2_MOUSE
 AC 035385; 16-OCT-2001 (Rel. 4.0, Created)
 DT 16-OCT-2001 (Rel. 4.0, Last sequence update)
 DT 28-FEB-2003 (Rel. 4.1, Last annotation update)
 DE Serine/threonine protein phosphatase with EF-hands-2 (EC 3.1.3.16)
 DE (PPEF-2).
 GN PPEF2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-EYE;
 RX MEDLINE=97471020; PubMed=9326663;
 RA Sherman P.M., Sun H., Macke J.P., Williams J., Smallwood P.M.,
 RA Nathans J.;
 RT "Identification and characterization of a conserved family of protein
 RT serine/threonine phosphatases homologous to Drosophila retinal
 RT degeneration C.";
 RT Proc. Natl. Acad. Sci. U.S.A. 94:11639-11644 (1997).
 RL -1- FUNCTION: MAY PLAY A ROLE IN PHOTOTRANSDUCTION. MAY
 CC DEPHOSPHORYLATE PHOTOACTIVATED RHODOPSIN. MAY FUNCTION AS A
 CC CALCIUM SENSING REGULATOR OF IONIC CURRENTS, ENERGY PRODUCTION OR
 CC SYNAPTIC TRANSMISSION.
 CC -1- CATALYTIC ACTIVITY: A phosphoprotein + H(2)O = a protein +
 CC phosphate.
 CC -1- COFACTOR: MAGNESIUM (BY SIMILARITY).
 CC -1- ENZYME REGULATION: ACTIVATED BY CALCIUM (BY SIMILARITY).
 CC -1- TISSUE SPECIFICITY: DETECTED IN RETINA, MORE SPECIFICALLY IN
 CC PHOTORECEPTORS.
 CC -1- SIMILARITY: BELONGS TO THE PPP FAMILY OF PHOSPHATASES.
 CC -1- SIMILARITY: Contains 1 IQ domain.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF023458; AAB82798.1; -.
 DR HSSP; P08129; 1FJM.
 DR MCD; MG1:1342304; Ppef2.
 DR InterPro; IPR002048; EF-hand.
 DR InterPro; IPR000048; IQ region.
 DR InterPro; IPR006842; M_ppef2ase.
 DR InterPro; IPR006185; S/T_phosphase.
 DR InterPro; IPR006186; T_thr_apah.
 DR Pfam; PF00036; ehand; 3.
 DR Pfam; PF00612; IQ; 1.
 DR Pfam; PF00149; Metallophos; 1.
 DR PRINTS; PR00114; STPHPTASE.
 DR ProDom; PD000012; EF-hand; 1.
 DR ProDom; PD000252; T_phase_apah; 2.
 DR SMART; SM00054; Efh; 3.
 DR SMART; SM00015; IQ; 1.
 DR SMART; SM00156; PP2Ac; 1.
 DR PROSITE; PS00018; EF_HAND; 3.
 DR PROSITE; PS00125; SER_THR_PHOSPHATASE; 1.
 DR PROSITE; PS50096; IQ; 1.
 KW Hydroxylase; Calcium-binding; Magnesium; Iron; Manganese; Repeat;
 KW Vision.
 FT DOMAIN 21 46 IQ.
 FT DOMAIN 128 544 CATALYTIC.
 FT DOMAIN 585 596 ANCESTRAL CALCIUM SITE.
 FT CA_BIND 669 680 EF-HAND 2 (POTENTIAL).
 FT CA_BIND 709 720 EF-HAND 3 (POTENTIAL).
 FT CA_BIND 179 179 IRON (BY SIMILARITY).
 FT METAL 179 179 IRON (BY SIMILARITY).
 FT METAL 181 181 IRON (BY SIMILARITY).
 FT METAL 208 208 IRON AND MANGANESE (BY SIMILARITY).
 FT METAL 240 240 MANGANESE (BY SIMILARITY).
 FT ACT_SITE 241 241 GENERAL ACID (BY SIMILARITY).
 FT METAL 292 292 MANGANESE (BY SIMILARITY).
 FT METAL 492 492 MANGANESE (BY SIMILARITY).
 SQ SEQUENCE 757 AA; 86644 MW; 821B3D1061AC00C8 CRC64;

Query Match 49.4%; Score 39; DB 1; Length 757;
 Best Local Similarity 50.0%; Pred. No. 26;
 Matches 6; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

QY 5 TXDVYQNIQYV 16
 DB 44 TWNIQFQSLFYNG 55

RESULT 13
 AMDS_ASPOR STANDARD; PRT; 545 AA.
 ID AMDS_ASPOR
 AC Q12559; 28-FEB-2003 (Rel. 4.1, Created)
 DT 28-FEB-2003 (Rel. 4.1, Last sequence update)
 DT 28-FEB-2003 (Rel. 4.1, Last annotation update)
 DE Acetamidase (EC 3.5.1.4).
 GN AMDS.
 OS Aspergillus oryzae.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 OC Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 OX NCBI_TaxID=5062;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RIB 40;
 RX MEDLINE=92104511; PubMed=1840550;
 RA Gomi K., Kitanoto K., Kumagai C.;
 RT "Cloning and molecular characterization of the acetamidase-encoding
 RT gene (amds) from Aspergillus oryzae.";
 RL Gene 108:91-98 (1991).
 CC -1- FUNCTION: Allows acetamide to be used as a sole carbon or
 CC nitrogen source.
 CC -1- CATALYTIC ACTIVITY: A monocarboxylic acid amide + H(2)O = a
 CC monocarboxylate + NH(3).

CC -1- SIMILARITY: BELONGS TO THE AMIDASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D10492; BA01373.1; -
 DR FIR: J06633; J06633.
 DR InterPro: IPR00120; Amidase.
 DR Pfam: PF01425; Amidase; 1.
 DR PROSITE: PS00571; AMIDASES; 1.
 DR KJ HydroLase.
 SQ SEQUENCE 545 AA; 60156 MW; 1DDFBCBCC7492C5B CRC64;
 Qy Query Match 46.7%; Score 38.5; DB 1; Length 545;
 Db Best Local Similarity 56.2%; Pred. No. 23;
 Matches 9; Conservative 4; Mismatches 2; Indels 1; Gaps 1;
 1 AEGSTXDYVQNIQYAG 16
 375 ADGSR-DYKHNINSG 389
 RESULT 14
 VP55_HUMAN STANDARD; PRT; 796 AA.
 ID VP55_HUMAN Q9H016; Q9H096; Q9H493; Q9H6J0; Q9N857; Q9NKG2; Q9NKG0;
 AC Q9NKG2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Vacuolar protein sorting 35 (Vesicle protein sorting 35) (VPS35)
 DE (Maternal embryonic 3).
 GN VPS35 OR MEM3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID:9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Lung;
 RX MEDLINE=20517226; PubMed=11062004;
 RA Edgar A.J., Polak J.M.;
 RT "Human homologues of yeast vacuolar protein sorting 29 and 35";
 RL Biochem. Biophys. Res. Commun. 277:622-630 (2000).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=20564209; PubMed=11123353;
 RA Zhang F., Yu L., Gao J., Fu Q., Dai F., Zhao Y., Zheng L., Zhao S.;
 RT "Cloning and characterization of human VPS35 and mouse Vps35 and
 RL mapping of VPS35 to human chromosome 16q13-q21";
 RL Genomics 70:253-257 (2000).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Colon;
 RX MEDLINE=20556005; PubMed=11102511;
 RA Rentfrew H.C., Sierra M.L., Balford R., Lesniak M.A., Barr V.A.,
 RT "Human orthologs of yeast vacuolar protein sorting proteins Vps26, 29,
 RL and 35: assembly into multimeric complexes";
 RL Mol. Biol. Cell 11:4105-4116 (2000).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Phaeochromocytoma;
 RA Peng Y., Li Y., Tu Y., Xu S., Han Z., Fu G., Chen Z.;
 RT "A novel gene expressed in human phaeochromocytoma";
 RL Submitted (Sep-1999) to the EMBL/GenBank/DBJ databases.
 [5]

RP SEQUENCE FROM N.A.
 RC TISSUE=ileal mucosa, placenta, and Testis/carcinoma;
 RA Itozaki T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Aotaka S., Yoshikawa Y.,
 RA Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K.,
 RA Hiro M., Ohmori Y., Ohashi M., Nishi T., Shibahara T., Tanaka T.,
 RA Shiraori A., Sudo H., Magatsuma M., Hosoi T., Kaku Y., Kodaira H.,
 RA Kondo H., Sugawara M., Takahashi M., Chiba Y., Ishida S., Murakawa K.,
 RA Ono Y., Takiguchi S., Matsumura H., Ishii S., Kawai Y., Saito K.,
 RA Matsumae S., Kimura K., Murakami K., Yamamoto T., Makemaru A.,
 RA Nakamura Y., Nagahara K., Masuno Y., Minomiyu K., Sasaki N.,
 RA Iwayanagi T.,
 RT "NEO human cDNA sequencing project";
 RL Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Testis;
 RX MEDLINE=21154917; PubMed=11230166;
 RA Wiernan S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
 RA Ansoerge W., Beecher M., Bloeker H., Baerends S., Blum H.,
 RA Lauber J., Duesterhoeft A., Beyer A., Koerner X., Strick N.,
 RA Wewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
 RA Wambutt R., Korn B., Klein M., Poustka A.;
 RT "Towards a catalog of human genes and proteins: sequencing and
 RT analysis of 500 novel complete protein coding human cDNAs";
 RL Genome Res. 11:422-435 (2001).
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain, and Prostate;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner K.H., Scheffer C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diachenko L., Mermisina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Scapleton M., Soares M.B., Bonaldo M.F., Casavate I.L., Schetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshylyuk S., Abramson R.D., Mullen S.J.,
 RA Raha S.S., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley J.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W.,
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko V., Bonifard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,
 RA Buterfield V.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
 RP SEQUENCE OF 469-796 FROM N.A.
 RC TISSUE=Leukemia;
 RA Zhou J., Yu W., Tang H., Mei G., Tsang Y.T.M., Bouck J., Gibbs R.A.,
 RA Margolin J.F.;
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RL -1- FUNCTION: May play a role in vesicular protein sorting, analogous
 CC to the yeast retromer proteins.
 CC -1- SUBUNIT: Part of a large multimer together with VPS26 and VPS29.
 CC Associates with SNX1 and probably also with SNX2.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated.
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in heart, brain,
 CC placenta, skeletal muscle, spleen, thymus, testis, ovary, small
 CC intestine, kidney and colon.
 CC -1- SIMILARITY: BELONGS TO THE VPS35 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 [5]

```

CC -----
DR EMBL AF191298; AAF02778.2; -
DR EMBL AF166382; AAG40619.1; -
DR EMBL AF175265; AAF89968.1; -
DR EMBL AF183418; AAG09687.1; -
DR EMBL AK001614; BAA91790.1; -
DR EMBL AK023650; BAA14626.1; ALT_INIT.
DR EMBL AK000395; BAA91137.1; ALT_INIT.
DR EMBL AL136888; CAB66822.1; -
DR EMBL BC002414; AAH02414.1; -
DR EMBL BC010362; AAH10362.1; -
DR EMBL AY007112; AAG01988.1; ALT_INIT.
DR Genew; HGNC:13487; VPS35.
DR MIM; 606931; -
DR GO; GO:0005829; Cytoosol; IDA.
DR GO; GO:0005515; F:protein binding activity; NMS.
DR GO; GO:0042147; P:retrograde (endosome to Golgi) transport; NMS.
DR InterPro; IPR005378; Vps35.
DR Pfam; PF03635; Vps35; 1.
DR Transport; Protein transport.
DR TRANSPORT; Protein transport.
FT CONFLICT 42 42 A -> S (IN REF. 6).
FT CONFLICT 160 160 I -> T (IN REF. 5; BAA14626).
FT CONFLICT 168 168 T -> P (IN REF. 3).
FT CONFLICT 453 453 S -> F (IN REF. 7; AAH10362).
FT CONFLICT 526 526 R -> G (IN REF. 5; BAA91790).
FT CONFLICT 694 694 K -> E (IN REF. 5; BAA91790).
FT CONFLICT 796 796 L -> H (IN REF. 5; BAA91137).
SQ SEQUENCE 796 AA; 91706 MW; 28D2DD1C6B20A0A CRC64;

Query Match
Best Local Similarity 47.1%; Score 38.5; DB 1; Length 796;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 AEG-STXDYVQNIQYAG 16
DB 86 AKGRKADLYELVOYAG 102

RESULT 15
VPS3_MOUSE STANDARD; PRT; 796 AA.
AC Q9EQH3; Q61123;
ID 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Vacuolar protein sorting 35 (Vesicle protein sorting 35) (Maternal-embryonic 3).
GN VPS35 OR MEM3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo; and Embryonic carcinoma;
RX MEDLINE=96327632; PubMed=8678978;
RA Hwang S.-Y., Benjamin L.E., Oh B., Rochstein J.L., Ackerman S.L.,
RA Bedington R.S.P., Solter D., Knowles B.B.,
RT "Genetic mapping and embryonic expression of a novel, maternally
transcribed gene Mem3."
RL Mamm. Genome 7:586-590 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564209; PubMed=11112353;
RA Zhang P., Yu L., Gao J., Fu Q., Dai F., Zhao Y., Zheng L., Zhao S.,
RT "Cloning and characterization of human VPS35 and mouse Vps35 and
mapping of VPS35 to human chromosome 16q13-q21."
RL Genomics 70:235-257 (2000).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast tumor;
RX MEDLINE=2388257; PubMed=12477932;

```

```

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Shaeffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein W.J., Uscin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.U., Abramson R.D., Mullan S.U.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz U., Myers R.W.,
RA Butlerfield Y.S.N., Kirywiniski M.I., Skalska U., Smaluk D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Maria M.A.,
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
CC - FUNCTION: May play a role in vesicular protein sorting, analogous
CC to the yeast retroviral proteins (by similarity).
CC - SUBUNIT: Part of a large multimer together with VPS26 and VPS29.
CC Associates with SNX1 and probably also with SNX2 (by similarity).
CC - SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated (by
CC similarity).
CC - TISSUE SPECIFICITY: Ubiquitous. Highly expressed in fat tissue,
CC testis, brain, kidney, thymus, liver and pancreas, and at lower
CC levels in heart, intestine and skeletal muscle. Detected in
CC oocytes, pre-implantation embryos and at E6.5-E12.5.
CC - CAUTION: BELONGS TO THE VPS35 FAMILY.
CC - CAUTION: Ref.1 sequence differs from that shown due to numerous
CC frame shifts.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL U47024; AAB18153.1; ALT_FRAME.
DR EMBL AF226323; AAG40621.1; -
DR EMBL BC0005469; AAH05469.1; -
DR EMBL BC006637; AAH06637.1; -
DR MGD; MGI:1890467; Vps35.
DR InterPro; IPR005378; Vps35.
DR Pfam; PF03635; Vps35; 1.
DR Transport; Protein transport.
DR TRANSPORT; Protein transport.
SQ SEQUENCE 796 AA; 91712 MW; 2ABD33811D641CC CRC64;

Query Match
Best Local Similarity 47.1%; Score 38.5; DB 1; Length 796;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 AEG-STXDYVQNIQYAG 16
DB 86 AKGRKADLYELVOYAG 102

```

Search completed: January 29, 2004, 14:44:51
 Job time : 11 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 14:42:13 ; Search time 35 Seconds

(without alignments)
117.967 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 AEGSTXDYQNTQYAG 16

Scoring table: BLOSUM62

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: SPTREMBL_23.*

1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_virus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	43	54.4	380	Q9CHN1	Q9CHN1 lactococcus
2	43	54.4	426	Q9L385	Q9L385 clostridium
3	42	53.2	380	Q97087	Q97087 streptococcus
4	42	53.2	380	Q8DP85	Q8DP85 streptococcus
5	42	53.2	470	Q8T0M4	Q8T0M4 methanococcus
6	41	51.9	241	Q8RT72	Q8RT72 streptomyces
7	41	51.9	562	Q8CD13	Q8CD13 mus musculus
8	41	51.9	2659	Q9VB07	Q9VB07 drosophila
9	40.5	51.3	209	Q68567	Q68567 xanthomonas
10	40.5	51.3	315	Q53010	Q53010 burkholderia
11	40.5	51.3	453	Q8R712	Q8R712 thermococcus
12	40	50.6	116	Q9UWZ4	Q9UWZ4 geannomyces
13	40	50.6	187	Q8XW00	Q8XW00 clostridium
14	40	50.6	295	Q9ZHR4	Q9ZHR4 azospirillum
15	40	50.6	338	Q8RED1	Q8RED1 fusobacterium
16	40	50.6	455	Q9JRX3	Q9JRX3 neisseria m

17	40	50.6	458	Q9UW11	Q9UW11 piromyces r
18	40	50.6	701	Q967D4	Q967D4 geodia cydo
19	40	50.6	724	Q21222	Q21222 caenorhabditis
20	40	50.6	764	Q8L5B2	Q8L5B2 chenopodium
21	40	50.6	16215	Q9NFS3	Q9NFS3 drosophila
22	40	50.6	18074	Q917U4	Q917U4 drosophila
23	39.5	50.0	60	Q8DP93	Q8DP93 streptococcus
24	39	49.4	142	Q8F0N8	Q8F0N8 streptococcus
25	39	49.4	191	Q9EW89	Q9EW89 streptomyces
26	39	49.4	248	Q8XRQ3	Q8XRQ3 raietonia s
27	39	49.4	301	Q8SRU2	Q8SRU2 encopallito
28	39	49.4	379	Q8BSV7	Q8BSV7 streptococcus
29	39	49.4	379	Q8B080	Q8B080 streptococcus
30	39	49.4	381	Q8DP53	Q8DP53 streptococcus
31	39	49.4	382	Q8RTW9	Q8RTW9 uncultured
32	39	49.4	623	Q8T838	Q8T838 listeria mo
33	39	49.4	683	Q9W6R4	Q9W6R4 fugu rubrip
34	39	49.4	771	Q9JH57	Q9JH57 human astro
35	39	49.4	771	Q9JH58	Q9JH58 human astro
36	39	49.4	771	Q67744	Q67744 human astro
37	39	49.4	771	Q9JH54	Q9JH54 human astro
38	39	49.4	771	Q9ICD1	Q9ICD1 human astro
39	39	49.4	771	Q9ICB9	Q9ICB9 human astro
40	39	49.4	771	Q9JH56	Q9JH56 human astro
41	39	49.4	771	Q9JH55	Q9JH55 human astro
42	39	49.4	771	Q9JH52	Q9JH52 human astro
43	39	49.4	771	Q9JH53	Q9JH53 human astro
44	39	49.4	771	Q9JH53	Q9JH53 human astro
45	39	49.4	829	P90754	P90754 caenorhabditis

ALIGNMENTS

RESULT 1

ID Q9CHN1 PRELIMINARY; PRT; 380 AA.

AC Q9CHN1; 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Glucose-1-phosphate adenyllyltransferase (EC 2.7.7.27).
GN GLUC OR LUG697.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403.
RX MEDLINE=21235186; PubMed=11337471; Jallion O., Malame K.,
RA Bolotin A., Wincker P., Manger S., Sorokin A.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.,
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
lactis ssp. lactis IL1403."
RL Genome Res. 11:721-753(2001).
DR EMBL: AE006303; AAK04795.1; -
DR InterPro: IPR005836; ADP_Glu_Pyroph.
DR InterPro: IPR001451; Hexapep_transf.
DR InterPro: IPR005835; NTP transferase.
DR Pfam: PF00132; hexapep_2.
DR Pfam: PF00483; NTP transferase; 1.
DR PROSITE: PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE: PS00810; ADP_GLC_PYROPHOSPH_3; 1.
KW Transferase; Nucleotidyltransferase; Complete proteome.
SQ SEQUENCE 380 AA; 42053 MW; 524B934EB96F68 CRC64;

Query Match 54.4%; Score 43; DB 16; Length 380;

Best Local Similarity 61.5%; Pred. No. 23;

Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYQNTQY 14
DB 100 EGSTXDYQNTQY 112


```

RESULT 2
Q9J385 PRELIMINARY; PRT; 426 AA.
ID Q9J385
AC Q9J385;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase.
GN GUGC.
OS Clostridium cellulosyticum.
OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxId=1521;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35319;
RA Ravaignani A., Jennert K.C.B., Steiner E., Gruenberg R., Jeffries J.R.,
RA Wilkinson S.R., Young D.I., Tidswell E.C., Brown D.P., Youngman P.J.,
RA Morris J.G., Young M.;
RT "Good directly controls the switch from acid to solvent production in
RT solvent-ferming clostridia.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ277601; CAB89882.1;
DR InterPro; IPR005836; ADP_Glu_pyrop.
DR InterPro; IPR005835; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR PROSITE; PS00809; ADP_GLC_PYROPHOSPH_2; 1.
DR Nucleotide; IPR00809; ADP_GLC_PYROPHOSPH_2; 1.
SQ SEQUENCE 426 AA; 47094 MW; E967B354E137A0D CRC64;

Query Match 54.4%; Score 43; DB 2; Length 426;
Best Local Similarity 61.5%; Pred. No. 26;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYQNIQY 14
DB 101 EGTANAVYQNIQY 113

RESULT 3
Q970S7 PRELIMINARY; PRT; 380 AA.
ID Q970S7
AC Q970S7;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase.
GN SP1122.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis W.R., Radune D.,
RA Holtzapple E., Kouri H., Wolf A.M., Uetrich T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL; AE007413; AAK75233.1;
DR TIGR; SP1122;
DR InterPro; IPR005836; ADP_Glu_pyrop.

```

```

DR InterPro; IPR005835; NTP_transferase.
DR Pfam; PF00483; NTP_transferase; 1.
DR PROSITE; PS00808; ADP_GLC_PYROPHOSPH_1; 1.
DR Nucleotide; IPR00809; ADP_GLC_PYROPHOSPH_2; 1.
SQ SEQUENCE 380 AA; 41490 MW; 23661026D2D04309 CRC64;

Query Match 53.2%; Score 42; DB 16; Length 380;
Best Local Similarity 53.8%; Pred. No. 35;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYQNIQY 14
DB 100 EGTSHAYQNIQY 112

RESULT 4
Q8DP85 PRELIMINARY; PRT; 380 AA.
ID Q8DP85
AC Q8DP85;
DT 01-MAR-2003 (TReMBLrel. 23, Created)
DT 01-MAR-2003 (TReMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Glucose-1-phosphate adenylyltransferase (EC 2.7.7.27).
GN GUGC OR SPRI030.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=171101;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Betrem S.T., Fritz L., Fu D.-D., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.C., Lee L.N., Leskowitz E.J., Lu J., Matsushima P.,
RA McArthur S.M., McNamee M., McNamee K., Mundy C.W., Nicas T.I.,
RA Norris P.H., O'Gara M., Peery R.B., Robertson G.I., Rokey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008475; PAK99834.1;
DR Nucleotide; IPR00809; ADP_GLC_PYROPHOSPH_2; 1.
SQ SEQUENCE 380 AA; 41572 MW; 2360E4E72DA1599 CRC64;

Query Match 53.2%; Score 42; DB 16; Length 380;
Best Local Similarity 53.8%; Pred. No. 35;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYQNIQY 14
DB 100 EGTSHAYQNIQY 112

RESULT 5
Q8TQW4 PRELIMINARY; PRT; 470 AA.
ID Q8TQW4
AC Q8TQW4;
DT 01-JUN-2002 (TReMBLrel. 21, Created)
DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
DT 01-MAR-2003 (TReMBLrel. 23, Last annotation update)
DE Transport protein.
GN MA516.
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanococci; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=2214;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RC MEDLINE=21929760; PubMed=11932228;
RA Galagan J.E., Nusbaum C., Roy A., Endrizzi M.G., MacDonald P.,

```

```

RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Altmor D., Brown A.,
RA Allen N., Naylor J., Stange-Thomann N., Deatellano K., Johnson R.,
RA Linton L., McEwan P., McKernan K., Talamas J., Tirrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Graham D.E., Grahame D.A., Guss A.M.,
RA Hederich R., Ingram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mubhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Unayam L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Bartell K.F., Jing H., Macario A.U.L., Paulsen I.,
RA Pritchett W., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.M., Birren B.,
RT "The genome of Methanosaeta acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AE010821; LAM04930.1; -.
DR InterPro; IPR007114; MFS.
DR InterPro; IPR005828; Sub_transporter.
DR Pfam; PF00083; sugar_tr; 1.
DR PROSITE; PS00850; MFS; 1.
DR Complete proteome.
SQ SEQUENCE 470 AA; 50953 MW; 4CD46B299C32B9D CRC64;

Query Match 53.2%; Score 42; DB 17; Length 470;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSTXDYVYONIQYAG 16
Db 390 GSTVBLNMGKAG 403

RESULT 6
Q9R172 PRELIMINARY; PRT; 241 AA.
ID Q9R172;
AC Q9R172;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Endo-1,4-beta-xylanase.
DE XINC OR SC00105 OR SCJ11.34C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinomycetales;
OC Streptomycetaceae; Streptomycetaceae; Streptomyces.
OC NCBI_Taxid=1902;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC Oliver K., Harris D.;
RT Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
RL [2]
RL SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC Bentley S.D., Parkhill J., Bartell B.G., Rajandream M.A.;
RT Submitted (Aug-1999) to the EMBL/Genbank/DBJ databases.
RL [3]
RL SEQUENCE FROM N.A.
RP STRAIN=A3(2);
RC MEDLINE=97000351; PubMed=8843436;
RA Kienast H., Hopwood D.A.;
RA Kienast H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb streptomyces coelicolor A3(2) chromosome."
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RN SEQUENCE FROM N.A.
RP STRAIN=A3(2); M145;
RC MEDLINE=2196410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kleser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares S., Taylor K.,

```

```

RA Warren T., Wietzorrek A., Woodward J., Bartell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939104; CAB52919.1; -.
DR HSSP; P09850; 1XNB.
DR InterPro; IPR001137; Glyco_hydro_11.
DR InterPro; IPR006311; Tat.
DR Pfam; PF00457; Glyco_hydro_11; 1.
DR PRINTS; PRO0911; GLYHYDRLASE11.
DR TIGRFAMs; TIGR01409; Tat_signal_seg; 1.
DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11_1; 1.
DR PROSITE; PS00777; GLYCOSYL_HYDROL_F11_2; 1.
DR GlycoSite; 376599DE51.7B78F CRC64;
SQ SEQUENCE 241 AA; 25845 MW; 376599DE51.7B78F CRC64;

Query Match 51.9%; Score 41; DB 16; Length 241;
Best Local Similarity 42.9%; Pred. No. 32;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYONIQY 14
Db 157 SDGSTYDLYOTTRY 170

RESULT 7
Q8CD13 PRELIMINARY; PRT; 562 AA.
ID Q8CD13;
AC Q8CD13;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein.
DE Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_Taxid=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISUB=Testis;
RC MEDLINE=22354683; PubMed=12466851;
RA The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573(2002).
DR EMBL; AK030006; BAC26729.1; -.
DR Hypothetical protein.
SQ SEQUENCE 562 AA; 63697 MW; 49202E45979C2299 CRC64;

Query Match 51.9%; Score 41; DB 11; Length 562;
Best Local Similarity 46.7%; Pred. No. 80;
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 2 EGSTXDYVYONIQ 13
Db 94 EGSTWDICQSIQ 105

RESULT 8
Q9VBUT PRELIMINARY; PRT; 2659 AA.
ID Q9VBUT;
AC Q9VBUT;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE CG11856 protein.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.

```

NCBI_TaxID=7227;
 OX [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=383KLEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Gelniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.M., Hoskins R.A., Galie R.F.,
 RA George A.R., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutcliffe G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazel R.G., Champe M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Panlock C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Bokoyva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fodde C., Gabriellian A.E., Garg N.S., Galibert W.M., Glaeser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris D., Mostrel A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Munz D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sider-Klamos I., Stimpson M., Stupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stepleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.R., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach U.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RI "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003751; AAF56430.1; -
 DR HSSP; P49792; IRP.
 DR FLYBASE; FBgn0039302; CG11856.
 DR InterPro; IPR000697; EVH1.
 DR InterPro; IPR000156; TIR.
 DR InterPro; IPR001440; TIR.
 DR InterPro; IPR001876; Znf_RangDP.
 DR Pfam; PF00638; Ran_BPL; 4.
 DR Pfam; PF00641; zif_RanBP; 2.
 DR SMART; SMO0160; RanBD; 4.
 DR SMART; SMO0547; znf_R32; 2.
 DR PROSITE; PS0196; RANB1; 4.
 DR PROSITE; PS01358; ZF_RANBP2_1; 2.
 DR PROSITE; PS0199; ZF_RANBP2_2; 2.
 DR PROSITE; PS0199; ZF_RANBP2_2; 2.
 SQ SEQUENCE 2659 AA; 292235 MW; BE406AD434D9BE03 CRC64;

Query Match 51.3%; Score 41; DB 5; Length 2659;
 Best Local Similarity 53.3%; Pred. No. 4.3e+02;
 Matches 8; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

DB 2 EGSTXDYQNIQYAG 16
 484 EGFORDYQNIQYAG 498

RESULT 9
 068567 PRELIMINARY; PRT; 209 AA.
 AC 068567;
 DT 01-AUG-1998 (TRENDEL. 07, Created)

DT 01-AUG-1998 (TRENDEL. 07, Last sequence update)
 DE 01-OCT-2000 (TRENDEL. 15, Last annotation update)
 DE Restriction endonuclease R.XbaI.
 GN XBAIR.
 OS Xanthomonas campestris.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
 OC Xanthomonadaceae; Xanthomonas.
 OC Xanthomonas.
 OX NCBI_TaxID=339;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=Badrii;
 RA Zhang B.-H., Wilson G.G.;
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF051092; AAC06983.1; -
 DR KRBAS6; 2126; XbaI.
 KW Endonuclease.
 SQ SEQUENCE 209 AA; 23834 MW; 6D3CB8CE3FA92061 CRC64;

Query Match 51.3%; Score 40.5; DB 2; Length 209;
 Best Local Similarity 50.0%; Pred. No. 34;
 Matches 10; Conservative 1; Mismatches 4; Indels 5; Gaps 1;

DB 2 EGSTXDYQNIQYAG 16
 49 EGSTXDYQNIQYAG 68

RESULT 10
 053010 PRELIMINARY; PRT; 315 AA.
 AC 053010;
 DT 01-NOV-1996 (TRENDEL. 01, Created)
 DT 01-NOV-1996 (TRENDEL. 01, Last sequence update)
 DT 01-DEC-2001 (TRENDEL. 19, Last annotation update)
 DE Hydroxyphenol-1, 2-dioxygenase.
 GN HADC.
 OS Burkholderia pickettii (Pseudomonas pickettii).
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
 OC Ralstoniaceae; Ralstonia.
 OX NCBI_TaxID=329;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=DRP0602;
 RA Hatuta T., Nakano O., Takizawa N., Kiyohara H.,
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 1-27 FROM N.A.
 RC STRAIN=DRP0602;
 RA Takizawa N., Yokoyama H., Yanagihara K., Hatuta T., Kiyohara H.;
 RT "A locus of *Pseudomonas pickettii* DRP0602, had that encodes 2,4,6-
 trichlorophenol-1,4-dechlorinase with hydroxylase activity, and
 RT hydroxylation of various chlorophenols by the enzyme.";
 RL J. Ferment. Biotech. 80:318-326(1995).
 DR EMBL; D86544; BAA13107.1; -
 DR HSSP; P00437; 3PCC.
 DR InterPro; IPR000627; Dioxygenase.
 DR Pfam; PF00775; Dioxygenase; 1.
 KW Dioxygenase.
 SQ SEQUENCE 315 AA; 34592 MW; B3BD3EF470406CA0 CRC64;

Query Match 51.3%; Score 40.5; DB 2; Length 315;
 Best Local Similarity 42.9%; Pred. No. 53;
 Matches 9; Conservative 4; Mismatches 1; Indels 7; Gaps 1;

DB 3 GSTXDYQNIQYAG 16
 143 GATIDWQADADGRYDQYAG 163

RESULT 11
 08R712 PRELIMINARY; PRT; 453 AA.
 AC 08R712;
 DT 08R712;

DT 01-JUN-2002 (TRENBLrel. 21, Created)
 DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Periplasmic protease.
 GN PRC3 OR TTE2617.
 OS Thermomicrobacterium tengcongensis.
 OC Bacteria; Firmicutes; Clostridia; Thermomicrobacteriales;
 OC Thermomicrobacteriaceae; Thermomicrobacter.
 OX NCBI_TaxID=119072;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=WB4 / JCM 11007;
 RX MEDLINE=21992816; PubMed=11997336;
 RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
 RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
 RA Tan H., Chen R., Wang J., Yu J., Yang H.;
 RT "A complete sequence of T. tengcongensis genome."
 RL Genome Res. 12:689-700(2002).
 DR EMBL; AE013202; AAM25737.1; -.
 DR MEROPS; S41.004; -.
 DR InterPro; IPR000225; Armadillo.
 DR InterPro; IPR001478; PDZ binding.
 DR InterPro; IPR002477; PG binding.
 DR InterPro; IPR004447; Protease_Cterm.
 DR InterPro; IPR003581; TSPc.
 DR Pfam; PF00595; PDZ; 1.
 DR SMART; SM00228; PDZ; 1.
 DR SMART; SM00245; TSPc; 1.
 DR TIGRfams; TIGR00225; prc; 1.
 DR PROSITE; PS01076; ARM_REPEAT; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Protease; Complete proteome.
 SQ SEQUENCE 453 AA; 4995 MW; 106503DE3FE5A74 CRC64;

Query Match 51.3%; Score 40.5; DB 16; Length 453;
 Best Local Similarity 64.3%; Pred. No. 78;
 Matches 9; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

QY 2 EGSTXDYVQNIQY 14
 Db 345 EGSTPDYVQNIQY 358

RESULT 12
 Q9TVZ4 PRELIMINARY; PRT; 116 AA.
 AC Q9TVZ4;
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
 DE Endo-1,4-beta-xylanase (EC 3.2.1.8) (Fragment).
 GN AXYL2.
 OS Gaeanomyces graminis.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariomycetes; incertae sedis; Magnaporthaceae; Gaeanomyces.
 OX NCBI_TaxID=29850;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Enami K.;
 RT "PCR-based characterization of fungal xylanase genes."
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: ENDOHYDROLYSIS OF 1,4-BETA-D-XYLANSIDIC
 CC LINKAGES IN XYLANS.
 CC -1- PATHWAY: XYLAN DEGRADATION.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY G (FAMILY 11 OF GLYCOSYL
 CC HYDROLASES).
 DR EMBL; AJ249160; CAB53513.1; -.
 DR HSSP; P09850; IXNB.
 DR InterPro; IPR001137; Glyco_hydro_11.
 DR Pfam; PF00457; Glyco_hydro_11; 1.
 DR PROSITE; PS00776; GLYCOSYL_HYDROL_F11.1; 1.
 KW Glycosidase; Hydrolase; Xylan degradation.

FT NON TER 1 1
 FT NON-TER 116 116
 SQ SEQUENCE 116 AA; 12791 MW; DA7BA4FBA6770B9E CRC64;
 Query Match 50.6%; Score 40; DB 3; Length 116;
 Best Local Similarity 42.9%; Pred. No. 22;
 Matches 6; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEGSTXDYVQNIQY 14
 Db 58 SDGFTDYVQNIQY 71

RESULT 13
 Q8XNW0 PRELIMINARY; PRT; 187 AA.
 AC Q8XNW0;
 DT 01-MAR-2002 (TRENBLrel. 20, Created)
 DT 01-MAR-2002 (TRENBLrel. 20, Last sequence update)
 DT 01-OCT-2002 (TRENBLrel. 22, Last annotation update)
 DE Hypothetical protein CPE0222.
 GN CPE0222.
 OS Clostridium perfringens.
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1502;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=13 / Type A;
 RX PubMed=11792842;
 RA Shida T., Ogasawara N., Hattori M., Kunara S., Hayashi H.;
 RA "Complete genome sequence of Clostridium perfringens, an anaerobic
 RT flesh-eater."
 RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
 DR EMBL; AP003185; BAB79928.1; -.
 DR InterPro; IPR005754; Sortase_fam; 1.
 DR TIGRfams; TIGR01076; Sortase_fam; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 187 AA; 20996 MW; A0C6CB3B5B13E2C CRC64;

Query Match 50.6%; Score 40; DB 16; Length 187;
 Best Local Similarity 72.7%; Pred. No. 37;
 Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEGSTXDYVQNIQY 11
 Db 77 AEGSTIDYVQNIQY 87

RESULT 14
 Q9ZHR4 PRELIMINARY; PRT; 295 AA.
 AC Q9ZHR4;
 DT 01-MAY-1999 (TRENBLrel. 10, Created)
 DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
 DE Cytochrome c oxidase Cyt c subunit.
 GN CYP.
 OS Aspergillus brasiliensis.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhodospirillales;
 OC Rhodospirillaceae; Aspergillum.
 OX NCBI_TaxID=192;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=57;
 RX MEDLINE=93009004; PubMed=9791120;
 RA Marchal K., Sun J., Keijzers V., Haaker H., Vanderleyden J.;
 RT "A cytochrome cbb3 (Cytochrome c) terminal oxidase in Aspergillum
 RT brasiliensis Sp7 supports microaerobic growth."
 RL J. Bacteriol. 180:5689-5696(1998).
 DR EMBL; AF054871; AAC72074.1; -.
 DR HSSP; P57736; IC60.

DR InterPro; IPR004678; CcoP.
 DR InterPro; IPR000345; CytC_heme_bind.
 DR InterPro; IPR003088; Cyt_C1.
 DR InterPro; IPR002329; Cyt_C1C.
 DR Pfam; PF00034; cytochrome c; 2.
 DR PRINTS; PRO0605; CYTOCHROME_C1.
 DR TIGRfams; TIGR00782; CCO2P; 1.
 DR PROSITE; PS00190; CYTOCHROME_C; 2.
 SQ SEQUENCE 295 AA; 31807 MW; 8F38D57A6E25F50F CRC64;

Query Match 50.6%; Score 40; DB 2; Length 295;
 Best Local Similarity 50.0%; Pred. No. 60;
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 EGSTXDYQNTQY 14
 |||:|:|:|:|:
 Db 147 GTTADYKTIQH 158

RESULT 15

Q8RED1 PRELIMINARY; PRT; 338 AA.
 AC Q8RED1;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein FN1177.
 GN FN1177.
 OS Fusobacterium nucleatum (subsp. nucleatum).
 OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;
 OC Fusobacterium.
 OX NCBI_TaxID=76856;
 RN [1]
 RP SEQUENCE FROM N. A.
 RC STRAIN-ATCC 25586;
 KK MEDLINE=2186394; PubMed=11889109;
 RA Kapatal V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,
 RA Bhattacharyya A., Bartman A., Gardner W., Grechkin G., Zhu L.,
 RA Vasileva O., Chu L., Kogan Y., Chaga O., Goldsman E., Bernal A.,
 RA Larsen N., D'Souza M., Mainas T., Pusch G., Haselkorn R.,
 RA Forstein M., Kyridis N., Overbeek R.;
 RT "Genome sequence and analysis of the oral bacterium Fusobacterium
 nucleatum strain ATCC 25586."
 RL J. Bacteriol. 184:2005-2018(2002).
 DR EMBL; AE010623; AAL95373.1; -.
 DR InterPro; IPR002729; DUF48.
 DR Pfam; PF01867; DUF48; 1.
 DR ProDom; PD008695; DUF48; 1.
 DR TIGRfams; TIGR00287; TIGR00287; 1.
 KM Hypothetical protein; Complete proteome.
 SQ SEQUENCE 338 AA; 40819 MW; CA00AA9225B885X2 CRC64;

Query Match 50.6%; Score 40; DB 16; Length 338;
 Best Local Similarity 38.5%; Pred. No. 70;
 Matches 5; Conservative 7; Mismatches 1; Indels 0; Gaps 0;

QY 2 EGSTXDYQNTQY 14
 |||:|:|:|:|:
 Db 118 EGASFNYRNLR 130

Search completed: January 29, 2004, 14:45:39
 Job time : 36 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 14:43:38 ; Search time 22 Seconds

(without alignments)
30,772 Million cell updates/sec

Title: US-09-737-297-3

Sequence: 1 AEGSTXDYVNIQYAG 16

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
6: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	41	51.9	191	1	US-08-044-621D-35
2	41	51.9	191	2	US-08-709-912-11
3	41	51.9	191	2	US-09-047-370-11
4	41	51.9	216	1	US-08-315-695-20
5	39	49.4	201	4	US-09-311-811C-21
6	39	49.4	313	4	US-09-551-826D-14
7	39	49.4	1260	4	US-09-328-352-6746
8	38.5	48.7	796	3	US-09-005-180A-1
9	38	48.1	200	1	US-07-744-570B-2
10	38	48.1	290	4	US-08-936-165A-462
11	38	48.1	291	4	US-09-322-981A-20970
12	38	48.1	365	4	US-09-328-352-7027
13	38	48.1	927	4	US-09-107-532A-4335
14	37	46.8	104	1	US-07-978-692-4
15	37	46.8	189	1	US-08-044-621D-33
16	37	46.8	189	1	US-08-709-912-12
17	37	46.8	189	2	US-09-047-370-12
18	37	46.8	200	3	US-08-275-826C-24
19	37	46.8	200	4	US-09-076-677-24
20	37	46.8	200	4	US-09-073-055-24
21	37	46.8	211	1	US-08-575-964-1
22	37	46.8	211	2	US-08-963-500-1
23	37	46.8	227	3	US-08-275-826C-31
24	37	46.8	227	4	US-09-076-677-31
25	37	46.8	227	4	US-08-073-055-31
26	37	46.8	365	4	US-08-198-452A-371
27	36	45.6	189	1	US-08-709-912-13

28	36	45.6	189	2	US-09-047-370-13	Sequence 13, Appl
29	36	45.6	431	2	US-09-328-352-7392	Sequence 7392, Ap
30	36	45.6	455	2	US-08-870-827-3	Sequence 3, Appl
31	36	45.6	455	3	US-09-317-179-3	Sequence 3, Appl
32	36	45.6	1455	3	US-08-840-062-5	Sequence 5, Appl
33	36	45.6	3135	3	US-08-323-170B-2	Sequence 2, Appl
34	36	45.6	3135	4	US-08-954-441-2	Sequence 2, Appl
35	35	44.3	98	4	US-09-252-991A-28763	Sequence 28763, A
36	35	44.3	174	4	US-09-198-452A-334	Sequence 334, App
37	35	44.3	184	1	US-08-088-633-6	Sequence 6, Appl
38	35	44.3	184	1	US-08-245-756-6	Sequence 6, Appl
39	35	44.3	184	1	US-08-441-750-6	Sequence 6, Appl
40	35	44.3	184	2	US-08-441-751-6	Sequence 6, Appl
41	35	44.3	184	5	PCT-US92-02521-6	Sequence 6, Appl
42	35	44.3	185	1	US-08-044-621D-36	Sequence 36, Appl
43	35	44.3	185	1	US-08-044-621D-37	Sequence 37, Appl
44	35	44.3	185	1	US-08-709-912-3	Sequence 3, Appl
45	35	44.3	185	1	US-08-709-912-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-044-621D-35
Sequence 35, Application US/08044621D
Patent No. 5405769
GENERAL INFORMATION:
APPLICANT: Warren W. Makarchuk
APPLICANT: Wing L. Sung
APPLICANT: Makoto Yaguchi
APPLICANT: Robert L. Campbell
APPLICANT: David R. Rose
TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS
TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE
NUMBER OF SEQUENCES: 37
CORRESPONDENCE ADDRESS:
ADDRESSEE: Gowling, Strathly & Henderson
STREET: Suite 2600, 160 Elgin Street
CITY: Ottawa
STATE: Ontario
COUNTRY: Canada
ZIP: K1P 1G3
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25 in., 360KB storage
COMPUTER: IBM PC
OPERATING SYSTEM: PC-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/044,621D
FILING DATE: April 8, 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Judy A. Estratt
REGISTRATION NUMBER: 34,076
REFERENCE/DOCKET NUMBER: 08-863796
TELECOMMUNICATION INFORMATION:
TELEPHONE: 613-786-0199
TELEFAX: 613-563-9869
TELEX:
INFORMATION FOR SEQ ID NO: 35:
SEQUENCE CHARACTERISTICS:
LENGTH: 191
TYPE: Amino Acid
STRANDEDNESS: No. 5405769 Relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
DESCRIPTION: No
HYPOTHETICAL: No
ANTI-SENSE: No

FRAGMENT TYPE: NO
ORIGINAL SOURCE:
ORGANISM: Streptomyces lividans
STRAIN: Streptomyces lividans, Xln C
IMMEDIATE SOURCE:
POSITION IN GENOME:
FEATURE:
PUBLICATION INFORMATION:
AUTHORS: Shareck, F., Roy, C., Yaguchi, M.,
AUTHORS: Morosoli, R. & Kluepfel, D.
TITLE:
JOURNAL: Gene
VOLUME: 107
ISSUE:
PAGES: 75-82
DATE: 1991
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-08-044-621D-35

Query Match 51.9%; Score 41; DB 1; Length 191;
Best Local Similarity 42.9%; Pred. No. 4;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQY 14

Db 107 SDGTYDYQYTRY 120

RESULT 2

US-08-709-912-11
Sequence 11, Application US/08709912
Patent No. 5759840
GENERAL INFORMATION:
APPLICANT: Sung Dr., Wing L
APPLICANT: Yaguchi Dr., Makoto
APPLICANT: Ishikawa Dr., Kazuhiko
TITLE OF INVENTION: Modification of Xylanase to Improve
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and
TITLE OF INVENTION: Thermostability
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto
STREET: 277 Park Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/709,912
FILING DATE: 09-SEP-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr., Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039.2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2982
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein

HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Streptomyces lividans
STRAIN: Xln C
PUBLICATION INFORMATION:
AUTHORS: Shareck, F
AUTHORS: Roy, C
AUTHORS: Yaguchi, M
AUTHORS: Morosoli, R
AUTHORS: Kluepfel, D
JOURNAL: Gene
VOLUME: 107
PAGES: 75-82
DATE: 1991
US-08-709-912-11

Query Match 51.9%; Score 41; DB 1; Length 191;
Best Local Similarity 42.9%; Pred. No. 4;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQY 14

Db 107 SDGTYDYQYTRY 120

RESULT 3

US-08-047-370-11
Sequence 11, Application US/09047370
Patent No. 5866408
GENERAL INFORMATION:
APPLICANT: Sung Dr., Wing L
APPLICANT: Yaguchi Dr., Makoto
APPLICANT: Ishikawa Dr., Kazuhiko
TITLE OF INVENTION: Modification of Xylanase to Improve
TITLE OF INVENTION: Thermophilicity, Alkalophilicity and
TITLE OF INVENTION: Thermostability
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fitzpatrick, Cella, Harper, and Scinto
STREET: 277 Park Ave.
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10172-0194
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/047,370
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/709,912
FILING DATE: 09-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Olsen Mr., Warren E
REGISTRATION NUMBER: 27290
REFERENCE/DOCKET NUMBER: 1039.2000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-2982
TELEFAX: (212) 758-2982
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 191 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO

ANTI-SENSE: NO
FRAGMENT TYPE: Internal
ORIGINAL SOURCE:
ORGANISM: Streptomyces lividans
STRAIN: Xln C
PUBLICATION INFORMATION:
AUTHORS: Shareck, F
AUTHORS: Roy, C
AUTHORS: Yaguchi, M
AUTHORS: Morosoli, R
AUTHORS: Kluepfel, D
JOURNAL: Gene
VOLUME: 107
PAGES: 75-82
DATE: 1991
US-09-047-370-11

Query Match 51.9%; Score 41; DB 2; Length 191;
Best Local Similarity 42.9%; Pred. No. 4;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYCNIOY 14
DB 107 SDGTYDIYQTTY 120

RESULT 4
US-08-315-695-20
Sequence 20, Application US/08315695
Patent No. 5591619
GENERAL INFORMATION:
APPLICANT: Li, Xin-Liang
APPLICANT: Ljungdahl, Lars G.
TITLE OF INVENTION: Aureobasidium Pullulans Xylanase, Gene
TITLE OF INVENTION: and Signal Sequence
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Greenlee and Winner, P.C.
STREET: 5370 Manhattan Circle, Suite 201
CITY: Boulder
STATE: CO
COUNTRY: US
ZIP: 80303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/315,695
FILING DATE: 30-SEP-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Winner, Ellen P.
REGISTRATION NUMBER: 28,547
REFERENCE/DOCKET NUMBER: 55-94
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 499-8080
TELEFAX: (303) 499-8089
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 216 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: Protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
US-08-315-695-20

Query Match 51.9%; Score 41; DB 1; Length 216;
Best Local Similarity 42.9%; Pred. No. 4.6;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
QY 1 AEGSTXDYVYCNIOY 14
DB 133 SDGTYDIYQTTY 146

RESULT 5
US-09-311-311C-21
Sequence 21, Application US/09311311C
Patent No. 6358728
GENERAL INFORMATION:
APPLICANT: Erikson, et al.
TITLE OF INVENTION: POLO BOX THERAPEUTIC COMPOSITIONS,
METHODS, AND USES THEREFOR
FILE REFERENCE: 1874/117
CURRENT APPLICATION NUMBER: US/09/311,311C
CURRENT FILING DATE: 1999-05-13
PRIOR APPLICATION NUMBER: US 60/085,296
PRIOR FILING DATE: 1998-05-13
NUMBER OF SEQ ID NOS: 27
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 21
LENGTH: 201
TYPE: PRT
ORGANISM: Drosophila melanogaster
FEATURE:
NAME/KEY: DOMAIN
LOCATION: (376)...(576)
OTHER INFORMATION: Polo protein C-terminal portion
US-09-311-311C-21

Query Match 49.4%; Score 39; DB 4; Length 201;
Best Local Similarity 53.8%; Pred. No. 10;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 3 GSTXDYVYCNIOYA 15
DB 176 GVSKDIYQXIRYA 188

RESULT 6
US-09-551-826D-14
Sequence 14, Application US/09551826D
Patent No. 6558939
GENERAL INFORMATION:
APPLICANT: No. 6558939regard-Madsen, Mads
APPLICANT: Ostergaard, Peter Rander
APPLICANT: Christensen, Claus Bo Voge
APPLICANT: Lassen, Soren Flenssted
TITLE OF INVENTION: No. 6558939el Proteases And Variants Thereof
FILE REFERENCE: 5665-200-US
CURRENT APPLICATION NUMBER: US/09/551,826D
CURRENT FILING DATE: 2000-04-17
NUMBER OF SEQ ID NOS: 14
SOFTWARE: Patent in version 3.2
SEQ ID NO 14
LENGTH: 313
TYPE: PRT
ORGANISM: Bacillus subtilis IS75
US-09-551-826D-14

Query Match 49.4%; Score 39; DB 4; Length 313;
Best Local Similarity 70.0%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 TXDYVYCNIOY 14
DB 300 TNDYVYCNIOY 309

RESULT 7
US-09-328-352-6746

Sequence 6746, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC09-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 6746
LENGTH: 1260
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-6746

Query Match 49.4%; Score 39; DB 4; Length 1260;
Best Local Similarity 58.3%; Pred. No. 89;
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 2 EGSTXDYQNIQ 13
DB 1080 EGFTREYVQSLQ 1091

RESULT 8
US-09-005-180A-1
Sequence 1, Application US/09005180A
Patent No. 6124446
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purvi
TITLE OF INVENTION: HUMAN VP35/MEM3-RELATED PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: PASCSE0 for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,180A
FILING DATE: Filed January 8, 1998
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0457 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 796 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: LUNGTUT08
CLONE: 2641812
US-09-005-180A-1

Query Match 48.7%; Score 38.5; DB 3; Length 796;
Best Local Similarity 47.1%; Pred. No. 64;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY 1 AEG-STXDYQNIQYAG 16
DB 86 AKGRXVADLYELVQYNG 102

RESULT 9
US-07-744-570B-2
Sequence 2, Application US/07744570B
Patent No. 5202249
GENERAL INFORMATION:
APPLICANT: Kluepfel, D.
APPLICANT: Morosoli, R.
APPLICANT: Shareck, F.
TITLE OF INVENTION: Xylanase for Biobleaching
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Michael J. Bradley
STREET: 1200 South 47th Street
CITY: Box Number 4023
CITY: Richmond
STATE: California
COUNTRY: United States
ZIP: 94804-0023
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44mb storage
COMPUTER: IBM
OPERATING SYSTEM: MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/744,570B
FILING DATE: 19910813
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 200 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single strand
TOPOLOGY: Circular
US-07-744-570B-2

Query Match 48.1%; Score 38; DB 1; Length 200;
Best Local Similarity 35.7%; Pred. No. 16;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQY 14
DB 116 SDGCTDYDYETTRY 129

RESULT 10
US-08-936-165A-462
Sequence 462, Application US/08936165A
Patent No. 6348582
GENERAL INFORMATION:
APPLICANT: Black, Michael
APPLICANT: Burnham, Martin
APPLICANT: Hodgson, John
APPLICANT: Knowles, David
APPLICANT: Lonetto, Michael
APPLICANT: Nicholas, Richard
APPLICANT: Pratt, Julie
APPLICANT: Reichard, Richard
APPLICANT: Rosenberg, Martin
APPLICANT: Ward, Judith
TITLE OF INVENTION: No. 6348582el Prokaryotic Polynucleotides,
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/936,165A
FILING DATE: 24-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/027,032
FILING DATE: 24-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Giumi, Edward R.
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX:
INFORMATION FOR SEQ ID NO: 462:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
US-08-936-165A-462

Query Match 48.1%; Score 38; DB 4; Length 290;
Best Local Similarity 40.0%; Pred. No. 24;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSTXDYQNIQYAG 16
DB 82 DGTIDLYEGIKETG 96

RESULT 11
US-09-252-991A-20970
Sequence 20970, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 20970
LENGTH: 291
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20970

Query Match 48.1%; Score 38; DB 4; Length 291;
Best Local Similarity 46.7%; Pred. No. 24;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQYIA 15
DB 175 SEGSIHQILQNVQQA 189

RESULT 12
US-09-328-352-7027
Sequence 7027, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:
APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
FILE REFERENCE: GTC99-039A
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 7027
LENGTH: 365
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-7027

Query Match 48.1%; Score 38; DB 4; Length 365;
Best Local Similarity 46.2%; Pred. No. 32;
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYQNIQY 14
DB 246 QGSTIDIFSNPOH 258

RESULT 13
US-09-107-532A-4335
Sequence 4335, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
CORRESPONDENCE ADDRESS:
NUMBER OF SEQUENCES: 7310
ADDRESS: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
City: Waltham
State: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Arinello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4335:
SEQUENCE CHARACTERISTICS:
LENGTH: 927 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1..927

SEQUENCE DESCRIPTION: SEQ ID NO: 4335

US-09-107-532A-4335

Query Match

48.1%; Score 38; DB 4; Length 927;

Best Local Similarity 70.0%; Pred. No. 95;

Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 STADYQNIQ 13

DB 681 SLEDIYQNIQ 690

RESULT 14

US-07-978-692-4

Sequence 4, Application US/07978692

Patent No. 5457045

GENERAL INFORMATION:

APPLICANT: Anker, Lisbeth

APPLICANT: Bisgaard-Frantzen, Henrik

APPLICANT: Halkier, Torben

TITLE OF INVENTION: ENZYMES WITH XYLANOXYLYTIC ACTIVITY

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESS: No. 54570450 No. 5457045disk of No. 5457045th America, Inc.

STREET: 405 Lexington Avenue, Suite 6200

CITY: New York

STATE: New York

COUNTRY: United States of America

ZIP: 10174-6201

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/978,692

FILING DATE: 19930202

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK91/00242

FILING DATE: 23-AUG-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: DK 0309/91

FILING DATE: 22-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/DK90/00220

FILING DATE: 24-AUG-1990

ATTORNEY/AGENT INFORMATION:

NAME: Lambiris, Elias J.

REGISTRATION NUMBER: 33728

REFERENCE/DOCKET NUMBER: 3399.204-US

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 104 amino acids

TYPE: amino acid

TOPOLOGY: linear

FRAGMENT TYPE: C-terminal

ORIGINAL SOURCE:

ORGANISM: Bacillus pumilus

STRAIN: DSM 6124

US-07-978-692-4

Query Match

46.8%; Score 37; DB 1; Length 104;

Best Local Similarity 38.5%; Pred. No. 11;

Matches 5; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

QY 1 AEGSTYDYQNIQ 13

DB 20 ADGGTYDIYETLR 32

RESULT 15

US-08-044-621D-33

Sequence 33, Application US/08044621D

Patent No. 5405769

GENERAL INFORMATION:

APPLICANT: Warren W. Makarchuk

APPLICANT: Wing L. Sung

APPLICANT: Makoto Yasuchi

APPLICANT: Robert L. Campbell

APPLICANT: David R. Rose

TITLE OF INVENTION: CONSTRUCTION OF THERMOSTABLE MUTANTS

TITLE OF INVENTION: OF A LOW MOLECULAR MASS XYLANASE

NUMBER OF SEQUENCES: 37

CORRESPONDENCE ADDRESS:

ADDRESS: Gowling, Scrathly & Henderson

STREET: Suite 2600, 160 Elgin Street

CITY: Ottawa

STATE: Ontario

COUNTRY: Canada

ZIP: K1P 1C3

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 in., 360KB storage

COMPUTER: IBM PC

OPERATING SYSTEM: PC-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/044,621D

FILING DATE: April 8, 1993

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Judy A. Erratt

REGISTRATION NUMBER: 34,076

REFERENCE/DOCKET NUMBER: 08-863796

TELECOMMUNICATION INFORMATION:

TELEPHONE: 613-786-0199

TELEFAX: 613-563-9869

TELEX:

INFORMATION FOR SEQ ID NO: 33:

SEQUENCE CHARACTERISTICS:

LENGTH: 189

TYPE: Amino Acid

STRANDEDNESS: No. 5405769 Relevant

TOPOLOGY: linear

MOLECULE TYPE:

DESCRIPTION: protein

HYPOTHETICAL: No

ANTI-SENSE: No

FRAGMENT TYPE: No

ORIGINAL SOURCE:

ORGANISM: Streptomyces sp.

STRAIN: Streptomyces sp. #36a

IMMEDIATE SOURCE:

POSITION IN GENOME:

FEATURE:

PUBLICATION INFORMATION:

AUTHORS: Nagashima M., Okumoto Y. & Okanishi M.

TITLE:

JOURNAL: Trends in Actinomycetologia

VOLUME:

ISSUE: 91-96

PAGES: 1989

DATE: 1989

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-044-621D-33

Query Match

46.8%; Score 37; DB 1; Length 189;

Mon Feb 2 10:57:13 2004

us-09-737-297-3.rat

Page 7

Best Local Similarity 35.7%; Pred. No. 22;
Matches 5; Conservative 5; Mismatches

4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVGNIOY 14
Db 107 SDGQYDLYKTRRY 120

Search completed: January 29, 2004, 14:46:46
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 29, 2004, 14:44:34 ; Search time 34 Seconds

(without alignments)
97.800 Million cell updates/sec

Title: US-09-737-297-3

Perfect score: 79

Sequence: 1 AEGSTXDYVQNIQYAG 16

Scoring table: BLOSUM62

Gapcp 10.0 , Gapext 0.5

Searched: 789580 seqs, 207824079 residues

Total number of hits satisfying chosen parameters: 789580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep.*
10: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep.*
14: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	ID	Description
1	77	97.5	16	US-09-737-297-3	Sequence 3, Appl
2	43	54.4	380	US-10-369-493-18356	Sequence 18356, A
3	41	51.9	191	US-10-307-441-11	Sequence 11, Appl
4	41	51.9	583	US-10-289-757-79	Sequence 79, Appl
5	40	50.6	456	US-09-815-242-10870	Sequence 10870, A
6	40	50.6	716	US-09-845-157-2	Sequence 2, Appl
7	40	50.6	724	US-10-369-493-5134	Sequence 5134, Ap
8	39	49.4	440	US-10-369-493-8931	Sequence 8931, Ap
9	38.5	48.7	813	US-09-764-888-197	Sequence 197, App
10	38	48.1	105	US-09-864-761-34387	Sequence 34387, A
11	38	48.1	290	US-09-935-960-462	Sequence 462, App
12	38	48.1	313	US-10-213-990-72	Sequence 72, Appl
13	38	48.1	497	US-10-358-917-14	Sequence 14, Appl
14	37.5	47.5	146	US-09-730-617-74	Sequence 74, Appl
15	37.5	47.5	149	US-09-730-617-68	Sequence 68, Appl

16	37.5	47.5	149	9	US-09-730-617-69	Sequence 69, Appl
17	37.5	47.5	149	9	US-09-730-617-70	Sequence 70, Appl
18	37.5	47.5	157	10	US-09-976-472-2	Sequence 2, Appl
19	37.5	47.5	157	12	US-10-139-947-2	Sequence 12, Appl
20	37.5	47.5	157	15	US-10-139-947-2	Sequence 17, Appl
21	37.5	47.5	157	15	US-10-139-947-2	Sequence 10, Appl
22	37.5	47.5	170	9	US-09-730-617-10	Sequence 77, Appl
23	37.5	47.5	170	9	US-09-730-617-77	Sequence 12, Appl
24	37	46.8	189	12	US-10-307-441-12	Sequence 18, Appl
25	37	46.8	211	12	US-10-237-386-18	Sequence 19, Appl
26	37	46.8	227	12	US-10-237-386-19	Sequence 35, Appl
27	37	46.8	240	12	US-10-237-386-35	Sequence 38, Appl
28	37	46.8	242	12	US-10-237-386-35	Sequence 41, Appl
29	37	46.8	301	10	US-09-738-626-6650	Sequence 6650, Ap
30	37	46.8	365	12	US-10-289-762-371	Sequence 371, App
31	37	46.8	423	12	US-10-369-493-2328	Sequence 2328, A
32	37	46.8	424	12	US-10-369-493-16517	Sequence 16517, A
33	37	46.8	447	12	US-10-312-273-241	Sequence 241, App
34	37	46.8	583	12	US-10-289-757-83	Sequence 83, Appl
35	37	46.8	728	12	US-10-287-274-300	Sequence 300, App
36	37	46.8	965	12	US-10-369-493-6940	Sequence 6940, Ap
37	36	45.6	189	12	US-10-307-441-13	Sequence 13, Appl
38	36	45.6	239	12	US-10-237-386-40	Sequence 40, Appl
39	36	45.6	240	12	US-10-237-386-42	Sequence 42, Appl
40	36	45.6	265	12	US-10-237-386-62	Sequence 62, Appl
41	36	45.6	455	10	US-09-986-682B-3	Sequence 3, Appl
42	36	45.6	647	12	US-10-094-749-2449	Sequence 2449, Ap
43	36	45.6	1456	10	US-09-870-759-95	Sequence 95, Appl
44	36	45.6	1456	12	US-09-751-708A-95	Sequence 95, Appl
45	36	45.6	1456	12	US-10-295-027-1217	Sequence 1217, Ap

ALIGNMENTS

RESULT 1
US-09-737-297-3
Sequence 3, Application US/09737297
Patent No. US20020072108A1
GENERAL INFORMATION:
APPLICANT: Berry, Mark
APPLICANT: Griffths, Allen
APPLICANT: Hall, Philip
APPLICANT: Laybourne-Parry, Johanna
APPLICANT: Mills, Sarah
TITLE OF INVENTION: Processes and Organisms for the Production of Antifreeze Prot
FILE REFERENCE: F3247
CURRENT APPLICATION NUMBER: US/09/737,297
PRIOR APPLICATION NUMBER: GB 9929696.4
PRIOR FILING DATE: 1999-12-15
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 3
LENGTH: 16
TYPE: PRT
ORGANISM: Marinomonas protea
FEATURES:
NAME/KEY: VARIANT
LOCATION: (6) ..(6)
OTHER INFORMATION: residue 6 is G or V
US-09-737-297-3
Query Match 97.5%; Score 77; DB 9; Length 16;
Best Local Similarity 100.0%; Pred. No. 7.9e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AEGSTXDYVQNIQYAG 16
DB 1 AEGSTXDYVQNIQYAG 16
RESULT 2

```

US-10-369-493-18356
; Sequence 18356, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18356
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Lactococcus lactis
US-10-369-493-18356

Query Match          54.4%; Score 43; DB 12; Length 380;
Best Local Similarity 61.5%; Pred. No. 21;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVYQNIQY 14
Db 100 EGTSHAVYQNIY 112

RESULT 3
US-10-307-441-11
; Sequence 11, Application US/10307441
; Publication No. US20030166236A1
; GENERAL INFORMATION:
; APPLICANT: Sung, Wing L.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Modified Xylanases Exhibiting Increased Thermophilicity
; TITLE OF INVENTION: and Alkalophilicity
; FILE REFERENCE: 027367-5006US
; CURRENT FILING DATE: 2002-12-02
; PRIOR APPLICATION NUMBER: PCT/CA01/00769
; PRIOR FILING DATE: 2001-05-31
; PRIOR FILING DATE: 2001-05-31
; PRIOR FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 60/213,803
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 191
; TYPE: PRT
; ORGANISM: Streptomyces lividans
US-10-307-441-11

Query Match          51.3%; Score 41; DB 12; Length 191;
Best Local Similarity 42.3%; Pred. No. 22;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDVYQNIQY 14
Db 107 SDGSTDYVYQNTY 120

RESULT 4
US-10-289-757-79
; Sequence 79, Application US/10289757
; Publication No. US20030180751A1
; GENERAL INFORMATION:
; APPLICANT: Demmer, Jerroen
; APPLICANT: Forster, Richard L.
; APPLICANT: Gibson, John Bryan

```

```

; APPLICANT: Shenk, Michael Andrew
; APPLICANT: No. US20030180751A1r1s, Geoffrey
; APPLICANT: Glenn, Matthew
; APPLICANT: Saulsbury, Keith Martin
; APPLICANT: Hall, Claire
; TITLE OF INVENTION: Compositions and methods for use
; TITLE OF INVENTION: grasses and methods for their use
; FILE REFERENCE: 11000.1061U
; CURRENT FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: 60/337,703
; PRIOR FILING DATE: 2001-11-07
; NUMBER OF SEQ ID NOS: 218
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 79
; LENGTH: 583
; TYPE: PRT
; ORGANISM: Lolium perenne
US-10-289-757-79

Query Match          51.9%; Score 41; DB 12; Length 583;
Best Local Similarity 53.8%; Pred. No. 75;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSTXDYVYQNIQY 14
Db 302 DNPTGDYVYQRIQY 314

RESULT 5
US-09-815-242-10870
; Sequence 10870, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10870
; LENGTH: 456
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10870

Query Match          50.6%; Score 40; DB 9; Length 456;
Best Local Similarity 46.7%; Pred. No. 85;
Matches 7; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

```

```
QY      2 EGSTXDYQNIQYAG 16
DB      255 EGSSFDVYHKADEVG 269

RESULT 6
US-09-845-157-2
; Sequence 2, Application US/09845157
; Patent No. US20020090618A1
; GENERAL INFORMATION:
; APPLICANT: Smith, J.
; TITLE OF INVENTION: Thermostable Reverse Transcriptases and Uses Thereof
; FILE REFERENCE: 0942.5040001
; CURRENT APPLICATION NUMBER: US/09/845,157
; CURRENT FILING DATE: 2001-05-01
; PRIOR APPLICATION NUMBER: US 60//207,196
; PRIOR FILING DATE: 2000-05-26
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 716
; TYPE: PRT
; ORGANISM: M-MuV reverse transcriptase gene
US-09-845-157-2

Query Match      50.6%; Score 40; DB 9; Length 716;
Best Local Similarity 53.3%; Pred. No. 1.4e+02;
Matches 8; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY      1 AEGSTXDYQNIQYA 15
DB      606 AEGKXJNVYNSRYA 620

RESULT 7
US-10-369-493-5134
; Sequence 5134, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5134
; LENGTH: 724
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-5134

Query Match      50.6%; Score 40; DB 12; Length 724;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY      1 AEGSTXDYQNI 12
DB      230 AEGSTDYVLTQL 241

RESULT 8
US-10-369-493-8931
; Sequence 8931, Application US/10369493
; Publication No. US2003023675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 8931
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Chloroflexus aurantiacus
US-10-369-493-8931

Query Match      49.4%; Score 39; DB 12; Length 440;
Best Local Similarity 53.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY      4 STPDYQNIQYAG 16
DB      143 NTEDVTGIEYRG 155

RESULT 9
US-09-764-898-197
; Sequence 197, Application US/09764898
; Patent No. US20020090673A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJ201
; CURRENT APPLICATION NUMBER: US/09/764,898
; CURRENT FILING DATE: 2001-01-17
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 311
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 197
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-898-197

Query Match      48.7%; Score 38.5; DB 9; Length 813;
Best Local Similarity 47.1%; Pred. No. 2.9e+02;
Matches 8; Conservative 4; Mismatches 4; Indels 1; Gaps 1;

QY      1 AEG-STXDYQNIQYAG 16
DB      103 AKGRKVDJELVQYAG 119

RESULT 10
US-09-864-761-34387
; Sequence 34387, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL R
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
```

PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 34387
LENGTH: 105
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC007869.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
OTHER INFORMATION: EST HUMAN HIT: BE06814.1, EVALUE 2.00e-36
OTHER INFORMATION: SWISSPROT HIT: P21414, EVALUE 2.00e-27
US-09-864-761-34387

Query Match 48.1%; Score 38; DB 9; Length 105;
Best Local Similarity 46.7%; Pred. No. 39;
Matches 7; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQYA 15
DB 41 SEGKTNTYTDSQYA 55

RESULT 11
US-09-939-980-462
Sequence 462, Application US/09939980
Patent No. US20020082234A1

GENERAL INFORMATION:
APPLICANT: Black, Michael
Burnham, Martin
Hodgson, John
Knowles, David
Lometto, Michael

Nicholas, Richard
Pratt, Julie
Reichard, Richard
Rosenberg, Martin
Ward, Judith
TITLE OF INVENTION: No. US20020082234A1 Prokaryotic Polynucleotides,
Polypeptides and Their Uses
NUMBER OF SEQUENCES: 534
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-0939
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/939,980
FILING DATE: 27-Aug-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/936,165
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Gimm, Edward R
REGISTRATION NUMBER: 38,891
REFERENCE/DOCKET NUMBER: P50549
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 462:
SEQUENCE CHARACTERISTICS:
LENGTH: 290 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
SEQUENCE DESCRIPTION: SEQ ID NO: 462:
US-09-939-980-462

Query Match 48.1%; Score 38; DB 9; Length 290;
Best Local Similarity 40.0%; Pred. No. 1,2e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSTXDYQNIQYNG 16
DB 82 DGTIDYEGIKETG 96

RESULT 12
US-10-213-990-72
Sequence 72, Application US/10213990
Patent No. US20030082595A1

GENERAL INFORMATION:
APPLICANT: Jiang, Bo
APPLICANT: Bussey, Howard
APPLICANT: Storms, Reg
APPLICANT: Roemer, Terry
TITLE OF INVENTION: NUCLEIC ACIDS OF ASPERGILLUS FUMIGATUS ENCODING INDUSTRIAL
FILE REFERENCE: 10182-019-999
CURRENT APPLICATION NUMBER: US/10/213,990
CURRENT FILING DATE: 2002-08-05
NUMBER OF SEQ ID NOS: 72
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 72
LENGTH: 313
TYPE: PRT

ORGANISM: Aspergillus
US-10-213-990-72

Query Match 48.1%; Score 38; DB 15; Length 313;
Best Local Similarity 46.2%; Pred. No. 1.3e+02;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 1 AEGSTXDYQNIQYAG 13
DB 143 SDGSTDYIEHQ 155

RESULT 13
US-10-358-917-14

Sequence 14, Application US/10358917
Publication No. US20030182687A1
GENERAL INFORMATION:
APPLICANT: Cheng, Qiong
APPLICANT: No. US20030182687A1ton, Kelley C.
APPLICANT: Tao, Luan
TITLE OF INVENTION: FUNCTIONALIZATION OF CAROTENOID COMPOUNDS
FILE REFERENCE: C11929 US NA
CURRENT APPLICATION NUMBER: US/10/358, 917
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: 60/355, 939
PRIOR FILING DATE: 2002-02-11
NUMBER OF SEQ ID NOS: 28
SOFTWARE: Microsoft Office 97
SEQ ID NO 14
LENGTH: 497
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-10-358-917-14

Query Match 48.1%; Score 38; DB 12; Length 497;
Best Local Similarity 40.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 EGSTXDYQNIQYAG 16
DB 94 DGTIDLYEGIKERG 108

RESULT 14
US-09-730-617-74

Sequence 74, Application US/09730617
Patent No. US20020068279A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E
APPLICANT: Prayaga, Sudhirdas K
APPLICANT: Shinkets, Richard A
APPLICANT: Rastelli, Luca
APPLICANT: Zernhusen, Bryan D
APPLICANT: Mezes, Peter S
TITLE OF INVENTION: NO. US20020068279A1el Proteins and Nucleic Acids Encoding the Sam
FILE REFERENCE: 15966-609
CURRENT APPLICATION NUMBER: US/09/730, 617
CURRENT FILING DATE: 2000-12-05
PRIOR APPLICATION NUMBER: 60/169, 056
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 60/169, 886
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/169, 866
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/170, 252
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 60/175, 740
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 74
LENGTH: 146
TYPE: PRT

ORGANISM: Homo sapiens
US-09-730-617-74

Query Match 47.5%; Score 37.5; DB 9; Length 146;
Best Local Similarity 53.3%; Pred. No. 69;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYQNIQYAG 16
DB 100 EGST-SVFGSVSTPG 113

RESULT 15
US-09-730-617-68

Sequence 68, Application US/09730617
Patent No. US20020068279A1
GENERAL INFORMATION:
APPLICANT: Burgess, Catherine E
APPLICANT: Prayaga, Sudhirdas K
APPLICANT: Shinkets, Richard A
APPLICANT: Rastelli, Luca
APPLICANT: Zernhusen, Bryan D
APPLICANT: Mezes, Peter S
TITLE OF INVENTION: NO. US20020068279A1el Proteins and Nucleic Acids Encoding the
FILE REFERENCE: 15966-609
CURRENT APPLICATION NUMBER: US/09/730, 617
CURRENT FILING DATE: 2000-12-05
PRIOR APPLICATION NUMBER: 60/169, 056
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 60/169, 886
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/169, 866
PRIOR FILING DATE: 1999-12-09
PRIOR APPLICATION NUMBER: 60/170, 252
PRIOR FILING DATE: 1999-12-10
PRIOR APPLICATION NUMBER: 60/175, 740
PRIOR FILING DATE: 2000-01-12
NUMBER OF SEQ ID NOS: 100
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 68
LENGTH: 149
TYPE: PRT
ORGANISM: Homo sapiens
US-09-730-617-68

Query Match 47.5%; Score 37.5; DB 9; Length 149;
Best Local Similarity 53.3%; Pred. No. 69;
Matches 8; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 2 EGSTXDYQNIQYAG 16
DB 103 EGST-SVFGSVSTPG 116

Search completed: January 29, 2004, 14:47:32
Job time : 34 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      January 29, 2004, 14:38:33 ; Search time 41 Seconds
              (without alignments)
              61.942 Million cell updates/sec
```

Title: US-09-737-297-3
Perfect score: 79
Sequence: 1 AEGSTXDVYQNIQYAG 16

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

```
Searched:      1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters:  1107863
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1:	A./genseq.16Jun03.*
2:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1980.DAT*
3:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1981.DAT*
4:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1982.DAT*
5:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1983.DAT*
6:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1984.DAT*
7:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1985.DAT*
8:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1986.DAT*
9:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1987.DAT*
10:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1988.DAT*
11:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1989.DAT*
12:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1990.DAT*
13:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1991.DAT*
14:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1992.DAT*
15:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1993.DAT*
16:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1994.DAT*
17:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1995.DAT*
18:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1996.DAT*
19:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1997.DAT*
20:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1998.DAT*
21:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA1999.DAT*
22:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA2000.DAT*
23:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA2001.DAT*
24:	/SIDS1./gcgcgata./genseq./geneseqp.-emb1./AA2003.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	77	97.5	16	22	AA884488
2	47	59.5	1164	22	ABG12516
3	43	54.4	380	23	ABBS4010
4	43	54.4	426	23	ABG31688
5	42	53.2	231	20	AA44183
6	42	53.2	380	24	ABU01542
7	42	53.2	387	23	ABG31652
8	41	51.9	191	13	AA860788
9	41	51.9	191	19	AA860261
					Modified xylanase
					N-terminal of anti
					Novel human diagn
					Lactococcus lactis
					Purine/pyrimidine
					Streptomyces clivi
					S. pneumoniae typ
					Purine/pyrimidine
					Xylanase C of Stre
					Modified xylanase

10	41	51.9	191	21	AA99675
11	41	51.9	191	21	AAE18447
12	41	51.9	200	14	AA832263
13	41	51.9	240	21	AA848555
14	41	51.9	240	21	AA018641
15	41	51.9	359	22	ABG31264
16	41	51.9	541	22	ACG08970
17	41	51.9	583	23	ABG59062
18	41	51.9	1560	22	ABG05639
19	41	51.9	2559	22	ABE84266
20	40	50.6	313	21	AAV15066
21	40	50.6	456	22	AAU35277
22	40	50.6	672	21	AAE12457
23	40	50.6	683	23	AAU49900
24	40	50.6	683	23	AAU49911
25	40	50.6	683	23	AAU49982
26	40	50.6	683	23	AAU49994
27	40	50.6	683	23	AAU49955
28	40	50.6	683	23	AAU49977
29	40	50.6	683	23	AAU49988
30	40	50.6	683	23	AAU49999
31	40	50.6	683	23	AAU75004
32	40	50.6	683	23	AAU75008
33	40	50.6	683	23	AAU75009
34	40	50.6	683	23	AAU75010
35	40	50.6	683	23	AAU75011
36	40	50.6	683	23	AAU75012
37	40	50.6	683	23	AAU75013
38	40	50.6	683	23	AAU75014
39	40	50.6	683	23	AAU75015
40	40	50.6	683	23	AAU75020
41	40	50.6	716	23	AAU75087
42	39.5	50.0	60	24	ABU01877
43	39	49.4	15	18	AAW01788
44	39	49.4	313	18	AAW04905
45	39	49.4	313	18	AAW16335

ALIGNMENTS

AC	AAB84468
XX	
XX	AAB84468; standard; peptide; 16 AA.
ID	AAB84468
RESULT	1
DT	05-SEP-2001 (first entry)
DE	N-terminal of anti-freeze peptide of Marinomonas protea.
XX	
KW	16S rRNA; anti-freeze protein; food additive; frozen vegetable; frozen confectionery.
OS	Marinomonas protea.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 6
XX	
PV	WO20014275-A2.
PD	21-JUN-2001.
XX	
PF	05-DEC-2000, 2000WO-EPI2396.
XX	
PR	15-DEC-1999; 99GB-0029696.
XX	
PA	(UNITL) UNILEVER PLC. (UNITL) UNILEVER NV.
PA	(HIND-) HINDUSTAN LEVER LTD.
PI	Berry MJ, Griffiths A, Hill PJ, Laybourne-Parry J, Mills SV

Rec'd, 8/19, 10/12-21/23

XX WP1: 2001-399120/42.
 DR
 XX
 PT Preparing anti-freeze peptides useful in frozen food products, e.g.
 PT frozen vegetables and confectionery, by culturing bacteria from aqueous
 PT low-temperature environment and extracting anti-freeze proteins from
 PT culture
 XX
 PS Claim 7, Page 9, 59pp; English.
 XX
 CC The present sequence represents the N-terminal of an anti-freeze
 CC peptide (AFP) of *Marinomonas protea*. *Marinomonas protea*. AFPs can be
 CC isolated from *M. protea*, using the method of the invention. The
 CC specification describes a method for producing AFPs. The method
 CC comprises collecting one or more samples of bacteria from an aqueous
 CC low-temperature environment, culturing the bacteria and extracting
 CC proteins from the samples, testing the proteins for anti-freeze
 CC properties, selecting proteins having anti-freeze properties and
 CC producing the selected protein for use as an AFP food additive. The
 CC method is useful for producing AFPs which are incorporated in food
 CC products, such as frozen vegetables and frozen confectionery such as
 CC ice-cream. AFPs are useful in frozen food products, such as vegetables,
 CC sauces, soups, snacks, dairy products and frozen confectionery, which
 CC includes sorbet, water-ice, granules, frozen fruit purees and
 CC milk-containing frozen products such as ice-cream, frozen yogurt or
 CC custards, sherbet and ice-milk.
 XX
 SQ Sequence 16 AA;
 Query Match 97.5%; Score 77; DB 22; Length 16;
 Best Local Similarity 100.0%; Pred. No. 7.8e-08;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 AEGSTXDYVYQNIQYAG 16
 DB 1 AEGSTXDYVYQNIQYAG 16
 RESULT 2
 ABG12516
 ID ABG12516 standard; Protein; 1164 AA.
 XX
 AC ABG12516;
 XX
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #12507.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 DR N-PSDB; AAS76703.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 20; SEQ ID No 42875; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIGO
 CC at ftp.wigo.in/pub/published_pat_sequences.
 XX
 SQ Sequence 1164 AA;
 Query Match 59.5%; Score 47; DB 22; Length 1164;
 Best Local Similarity 53.3%; Pred. No. 9.3;
 Matches 8; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 OY 2 EGSTXDYVYQNIQYAG 16
 DB 13 DGSSTAEFYQRIKYL 27
 RESULT 3
 ABB54010
 ID ABB54010 standard; Protein; 380 AA.
 XX
 AC ABB54010;
 XX
 DT 16-MAY-2002 (first entry)
 DE Lactococcus lactis protein glc.
 XX
 KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
 XX
 OS Lactococcus lactis IL1403.
 PN FR2807446-A1.
 XX
 PD 12-OCT-2001.
 XX
 PF 11-APR-2000; 2000FR-0004630.
 XX
 PR 11-APR-2000; 2000FR-0004630.
 XX
 PA (INRG) INRA INST NAT RECH AGRONOMIQUE.
 XX
 PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
 DR WP1: 2002-043418/06.
 XX
 PT New nucleotide sequence useful in the identification of Lactococcus
 PT lactis and related species -
 XX
 PS Claim 6; SEQ ID No 712; 2504pp; French.
 XX
 CC The present invention is related to a Lactococcus lactis nucleotide
 CC sequence (ABR90521) and related proteins (ABB53300-ABB55621). The
 CC nucleic acid sequence is useful in the detection and/or amplification of
 CC nucleic acid sequence, particularly to identify Lactococcus lactis or

CC related species. The proteins of the invention are useful for the
 CC biosynthesis or biodegradation of a composition of interest. The
 CC invention helps research in lactic bacteria, particularly useful in the
 CC production of yogurt and cheese.
 CC Note: The sequence data for this patent is based on equivalent patent
 CC WO20017734 (published 18-OCT-2001) which is available in electronic
 CC format directly from WIPO at http://wipo.int/pub/publ/published_pcc_sequences.
 CC XX

SQ Sequence 380 AA;

Query Match 54.4%; Score 43; DB 23; Length 380;
 Best Local Similarity 61.5%; Pred. No. 13;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVONTQY 14
 Db 100 EGTSNAYVONTISY 112

RESULT 4
 ABG91688
 ID ABG91688 standard; Protein; 426 AA.
 AC ABG91688;
 DT 18-NOV-2002 (first entry)
 DE Purine/pyrimidine triphosphate type nucleotidyltransferase #273.
 XX
 XX Nucleotidyltransferase; enzyme; active site engineering;
 KM alpha-D-glucopyranosyl phosphate thymidyltransferase; Epi;
 KM substrate specificity; nucleotide sugar;
 KM glycosylated bioactive natural product.
 XX
 OS Clostridium cellulolyticum.
 XX
 PN MO200248331-A2.
 PD 20-JUN-2002.
 PF 13-DEC-2001; 2001MO-US47953.
 PR 13-DEC-2000; 2000US-254927P.
 XX
 XX (SLOK) SLOAN KETTERING INST CANCER RES.
 PA
 P1 Thorson JS, Niklov DB;
 DR WPI; 2002-608282/65.
 XX
 PT Nucleotidyltransferase mutated at one or more amino acids, useful in
 PT the synthesis of nucleotide sugars -
 XX
 PS Claim 3; Page -; 182pp; English.
 XX
 CC The invention relates to a Nucleotidyltransferase mutated at one or
 CC more amino acids selected from V173, G147, W224, N112, G175, D111, E162,
 CC T201, I200, E199, R195, L897, L109, Y146 or Y177 (with reference to
 CC the Salmonella enterica rnaA-encoded alpha-D-glucopyranosyl phosphate
 CC thymidyltransferase, Epi, enzyme appearing as ABG91798). The mutations
 CC alter the substrate specificity of the enzymes. The mutants and methods
 CC involving them are used in the synthesis of nucleotide sugars for
 CC altering nucleotidyltransferase substrate specificity. The
 CC nucleotidyltransferase exhibits different substrate specificity for
 CC GTP, CTP, TTP and ATP than a non-mutated nucleotidyltransferase.
 CC The mutant may also exhibit a high degree of sequence identity to
 CC Salmonella enterica LT2 alpha-D-glucopyranosyl phosphate
 CC thymidyltransferase (Epi) and can convert a wide variety of phosphates.
 CC The mutants can be exploited in the biosynthesis of glycosylated
 CC bioactive natural products of pharmaceutical use. The present
 CC sequence is a nucleotidyltransferase exhibiting a high degree of
 CC sequence identity to Salmonella enterica LT2 alpha-D-glucopyranosyl
 CC phosphate thymidyltransferase (Epi).

CC Note: The present sequence is not displayed in the specification but was
 CC obtained from Genbank.
 CC XX

SQ Sequence 426 AA;

Query Match 54.4%; Score 43; DB 23; Length 426;
 Best Local Similarity 61.5%; Pred. No. 15;
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 EGSTXDYVONTQY 14
 Db 101 KGTANAYVONTQY 113

RESULT 5
 AAY44183
 ID AAY44183 standard; Protein; 231 AA.
 AC AAY44183;
 DT 01-FEB-2000 (first entry)
 DE Streptomyces olivaceoviridis xylanase (XynG) protein.
 XX
 XX Xylanase; plasmid; expression; E.coli; xylO-oligosaccharide; xylan;
 KM pulp-bleaching.
 XX
 OS Streptococcus olivaceoviridis.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..40
 FT /label= signal_peptide
 FT 41..399
 FT Protein /label= mature protein
 FT /note= "the mature protein and its coding sequence
 FT are claimed"

XX
 PN JP1266873-A.
 PD 05-OCT-1999.
 PF 20-MAR-1998; 98JP-0090702.
 PR 20-MAR-1998; 98JP-0090702.
 XX
 XX (NOR) NORINSUISANSO SHOKUHIN SOGO.
 PA (SEIB-) SEIBUTSUKI TOKUKEI SANGYO GIUTSU.
 XX
 DR WPI; 1999-613780/53.
 DR N-PSDB; AA228864.
 XX
 PT A xylanase gene, contiguous with a vector and a transformant - used for
 PT pulp-bleaching
 XX
 PS Example 1; Page 5-6; 10pp; Japanese.
 XX
 CC This sequence corresponds to the xynG xylanase Streptomyces
 CC olivaceoviridis. The coding sequence can be inserted into the plasmid
 CC pGB60 to generate plasmid pGB60-16713 for expression in e.g. E.coli.
 CC The xylanase is useful for the preparation of xylO-oligosaccharide from
 CC xylan and for pulp-bleaching.
 XX

SQ Sequence 231 AA;

Query Match 53.2%; Score 42; DB 20; Length 231;
 Best Local Similarity 50.0%; Pred. No. 11;
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVONTQY 14
 Db 149 SDGTYDVYQTRRY 162

RESULT 6
 ID AB001542 standard; Protein; 380 AA.
 AC AB001542;
 XX
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #118.
 XX
 XX Bacterial meningitis; pneumonia; sepsis; otitis media;
 XX ear infection; antiinflammatory; antibacterial; immunostimulant;
 XX auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 XX
 XX WO200277021-A2.
 PN 03-OCT-2002.
 PD
 XX
 XX 27-MAR-2002; 2002WO-IB02163.
 PF
 XX 27-MAR-2001; 2001GB-0007658.
 PR
 XX
 XX (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Massignani V, Tettelin H, Fraser C;
 PI WPI; 2003-040579/03.
 DR N-PSDB; ABX06830.
 XX
 XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection
 PS
 PS Claim 1; SEQ ID NO 2236; 56bp; English.
 XX
 XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS6644. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. The methods are useful for
 CC vaccines, diagnostics and antibiotics. The present sequence is one of
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pcr_sequences.
 XX
 XX Sequence 380 AA.
 SQ

```

Query Match Score 42; DB 24; Length 380;
Best Local Similarity 53.8%; Pred. No. 21;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY      2  EGSITDYVONIQT 14
      |||: |||
DB      100  EGSITSHYVONIDY 112

RESULT 7
ID      ABG91652
       ABG91652 standard; Protein; 387 AA.
AC      ABG91652;
DT      18-NOV-2002 (first entry)
DE      Purine/pyrimidine triphosphate type nucleotidyltransferase #237.
KM      Nucleotidyltransferase; enzyme; active site engineering;
KM      alpha-D-glucopyranosyl phosphate thymidyltransferase; Ep;
KM      substrate specificity; nucleotide sugar;
KM      glycosylated bioactive natural product.
OS      Geobacillus stearthermophilus.
XX      MO200248331-A2.
XX      20-JUN-2002.
XX      13-DEC-2001; 2001WO-US47953.
XX      13-DEC-2000; 2000US-254927P.
XX      (SLOC ) SLOAN KETERING INST CANCER RES.
PA      Thorson JS, Nikilov DB;
PI      WPI; 2002-608282/65.
DR      Nucleotidyltransferase mutated at one or more amino acids, useful in
PT      the synthesis of nucleotide sugars -
XX      Claim 3; Page -; 182pp; English.
PS      The invention relates to a Nucleotidyltransferase mutated at one or
XX      more amino acids selected from V173, G147, W224, N112, G175, D111, E162,
CC      T201, I200, E199, R195, L897, L109, Y146 or Y177 (with reference to
CC      the Salmonella enterica rnaA-encoded alpha-D-glucopyranosyl phosphate
CC      thymidyltransferase, Ep, enzyme appearing as ABG91798). The mutations
CC      alter the substrate specificity of the enzymes. The mutants and methods
CC      involving them are used in the synthesis of nucleotide sugars for
CC      altering nucleotidyltransferase substrate specificity. The
CC      nucleotidyltransferase exhibits different substrate specificity for
CC      GTP, CTP, UTP and ATP than a non-mutated nucleotidyltransferase.
CC      The mutant may also exhibit a high degree of sequence identity to
CC      Salmonella enterica LT2 alpha-D-glucopyranosyl phosphate
CC      thymidyltransferase (Ep) and can convert a wide variety of phosphates.
CC      The mutants can be exploited in the biosynthesis of glycosylated
CC      bioactive natural products of pharmacological use. The present
CC      sequence is a nucleotidyltransferase exhibiting a high degree of
CC      sequence identity to Salmonella enterica LT2 alpha-D-glucopyranosyl
CC      phosphate thymidyltransferase (Ep).
CC      Note: The present sequence is not displayed in the specification but was
CC      obtained from Genbank.
XX      SQ      Sequence      387 AA;
XX
XX      Query Match      53.2%; Score 42; DB 23; Length 387;
XX      Best Local Similarity 51.5%; Pred. No. 21;
XX      Matches      8; Conservative      1; Mismatches      4; Indels      0; Gaps      0;
XX      2  EGSITDYVONIQT 14

```

Db 100 EGTANAVYQNIY 112

RESULT 8

AAW60738 standard; protein; 191 AA.

AAW60738;

02-SEP-1998 (first entry)

Xylanase C of Streptomyces lividans.

Family 11 xylanase; improve; thermophilicity; alkalophilicity; thermotolerance; bleach; wood pulp; processing; wheat; maize; digestibility-improving animal feed additive; starch production; mutant.

Streptomyces lividans.

EP828002-A2.

11-MAR-1998.

05-SEP-1997; 97EP-0115412.

09-SEP-1996; 96US-0709912.

(CANADA) NAT RES COUNCIL CANADA.

(NARE-) NAT RES COUNCIL.

Ishikawa K, Sung WL, Yaguchi M;

WPI; 1998-161100/15.

Modified xylanase enzymes - useful for improving wood pulp

bleaching, etc.

Disclosure; Pages 42-43; 84pp; English.

AAW60728-44 represent family 11 xylanases. The specification describes a method for modifying a family 11 xylanase to improve its thermophilicity, alkalophilicity and or thermotolerance. This method comprises modification of amino acids 10, 27 or 29 of Trichoderma reesei xylanase II or corresponding aligned amino acid sequences in the xylanase, replacement of one or more amino acid sequences in the N-terminal region with corresponding aligned sequences from another family 11 xylanase to form a chimeric xylanase and/or upstream extension of the N terminus by addition of up to 10 amino acids. The modified xylanases are useful for improving the bleachability of wood pulp by treatment at 55-75 degrees Celsius and pH 7.5-9.0 for 5-180 minutes. They might also be useful as digestibility-improving animal feed additives. They might also be useful in the processing of wheat or maize for starch production.

Sequence 191 AA;

Query Match 51.9%; Score 41; DB 19; Length 191;

Best Local Similarity 42.9%; Pred. No. 13;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

1 AEGSTXDVYQNIY 14

107 SDGTYDIYQITRY 120

RESULT 9

AAW60281 standard; protein; 191 AA.

AAW60281;

02-SEP-1998 (first entry)

Modified xylanase C of Streptomyces lividans.

Family 11 xylanase; improve; thermophilicity; alkalophilicity; thermotolerance; bleach; wood pulp; processing; wheat; maize; digestibility-improving animal feed additive; starch production; mutant.

Synthetic.

Streptomyces lividans.

Key Location/Qualifiers

Misc-difference 15 /label= Y15X

EP828002-A2.

11-MAR-1998.

05-SEP-1997; 97EP-0115412.

09-SEP-1996; 96US-0709912.

(CANADA) NAT RES COUNCIL CANADA.

(NARE-) NAT RES COUNCIL.

Ishikawa K, Sung WL, Yaguchi M;

WPI; 1998-161100/15.

Modified xylanase enzymes - useful for improving wood pulp

bleaching, etc.

Claim 5; Page -; 84pp; English.

The present sequence represents a modified xylanase B of Streptomyces lividans. The specification describes a method for modifying a family 11 xylanase to improve its thermophilicity, alkalophilicity and or thermotolerance. This method comprises modification of amino acids 10, 14, 27 or 29 of Trichoderma reesei xylanase II or the corresponding aligned amino acid sequences in the N-terminal region with corresponding aligned sequences from another family 11 xylanase to form a chimeric xylanase and/or upstream extension of the N terminus by addition of up to 10 amino acids. The modified xylanases are useful for improving the bleachability of wood pulp by treatment at 55-75 degrees Celsius and pH 7.5-9.0 for 5-180 minutes. They might also be useful as digestibility-improving animal feed additives. They might also be useful in the processing of wheat or maize for starch production. note: this sequence does not appear in the specification; it was created using information provided.

Sequence 191 AA;

Query Match 51.9%; Score 41; DB 19; Length 191;

Best Local Similarity 42.9%; Pred. No. 13;

Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

1 AEGSTXDVYQNIY 14

107 SDGTYDIYQITRY 120

RESULT 10

AAV99675 standard; protein; 191 AA.

AAV99675;

28-SEP-2000 (first entry)

S. lividans xylanase, Xyn C.

KW Xylanase; animal feed; digestion efficiency; thermostable;
 KW feed pelleting; enzyme; Xyn A; Xyn B; Xyn; Xyn C; Xyn I; Xyn II.
 OS Streptomyces lividans.
 PN WO200029587-A1.
 XX 25-MAY-2000.
 PD 16-NOV-1999; 99WO-CA01093.
 PF 16-NOV-1999; 98US-0108504.
 XX 16-NOV-1998; 98US-0108504.
 PA (IOGE-) IOGEN CORP.
 XX Sung WL, Tolan JS;
 PI WPI; 2000-387799/33.
 DR WPI; 2000-387799/33.
 XX Thermostable xylanases useful for preparing animal feeds especially
 PT poultry or swine feed, exhibits optimal activity under physiological
 PT conditions
 XX
 PS Disclosure; Fig 1; 86pp; English.
 XX Xylanase enzymes are added to animal feeds to increase the efficiency of
 CC digestion and assimilation of nutrients. Xylanases are preferentially
 CC added during the feed pelleting process. To survive the pelleting
 CC process and to have optimum activity in the animal, the xylanase needs to
 CC have high thermostability, with optimum activity at physiological pH and
 CC temperature. The present sequence, xylanase Xyn C, from Streptomyces
 CC lividans, is a xylanase Family 11 member. The xylanases of Family
 CC 11 have several properties suitable for feed applications, however, they
 CC lack the thermostability required to survive food pelleting. The present
 CC sequence was used to identify non-conserved residues in Family 11
 CC xylanases which could be mutated to introduce desirable properties e.g.
 CC thermostability. As a result various thermostable xylanases were
 CC identified (AAV99683, AAV99684, AAV99685, AAV99735 and
 CC AAV99736) which would be useful for animal feeds, especially poultry and
 CC swine feed.
 XX
 SQ Sequence 191 AA;
 XX
 Query Match 51.9%; Score 41; DB 21; Length 191;
 Best Local Similarity 42.9%; Pred. No. 13;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAGSTYDVYQNIQY 14
 DB 107 SDGTYDIYQTRRY 120
 XX
 RESULT 11
 AAE18447
 ID AAE18447 standard; Protein; 191 AA.
 XX
 AC AAE18447;
 XX
 DT 16-MAY-2002 (first entry)
 XX
 DE Streptomyces lividans xylanase, Xyn C.
 XX Modified xylanase; thermostability; alkalophilicity; industrial process;
 KW pulp manufacture; poultry; swine feed; enzyme; Xyn C.
 XX
 OS Streptomyces lividans.
 XX
 PN WO200192487-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 31-MAY-2001; 2001WO-CA00769.
 XX

PR 31-MAY-2000; 2000US-213803P.
 XX
 XX (CANA) NAT RES COUNCIL CANADA.
 XX
 PI Sung WL;
 XX
 DR WPI; 2002-171435/22.
 XX
 XX Modified xylanase exhibiting increased thermostability and
 PT alkalophilicity useful for industrial processing e.g. for pulp
 PT manufacturing
 XX
 PS Disclosure; Page 77-78; 109pp; English.
 XX
 CC The present invention relates to a modified xylanase exhibiting increased
 CC thermostability and alkalophilicity. Modified xylanase is useful in also
 CC industrial process such as pulp manufacturing. Modified xylanase is also
 CC useful for bleaching of pulp, processing of precision devices and
 CC improving digestibility of poultry and swine feed. Modified xylanase has
 CC improved performance at conditions of high temperature and pH and
 CC exhibits improved thermostability and/or alkalophilicity in comparison to
 CC corresponding native xylanase. The present sequence is
 CC Streptomyces lividans xylanase, Xyn C.
 XX
 SQ Sequence 191 AA;
 XX
 Query Match 51.9%; Score 41; DB 23; Length 191;
 Best Local Similarity 42.9%; Pred. No. 13;
 Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
 QY 1 AAGSTYDVYQNIQY 14
 DB 107 SDGTYDIYQTRRY 120
 XX
 RESULT 12
 AAR32263
 ID AAR32263 standard; Protein; 200 AA.
 XX
 AC AAR32263;
 XX
 DT 25-MAR-2003 (updated)
 DT 15-JUN-1993 (first entry)
 XX
 DE Endo-xylanase from Streptomyces.
 XX
 XX Xinc; xyloisidic linkages; pulp; lignocellulose; brightening;
 XX viscosity; swelling.
 XX
 OS Streptomyces lividans.
 XX
 PN WO9303155-A1.
 XX
 PD 18-FEB-1993.
 XX
 PF 10-AUG-1992; 92WO-CA00349.
 XX
 PR 08-AUG-1991; 91CA-2048322.
 XX
 PA (FRAP-) FRAPPIER ARMAND INST.
 XX
 PI Kluepfel D, Morosoli R, Shareck F;
 XX
 DR WPI; 1993-076512/09.
 XX
 DR N-PADB; AAQ36563.
 XX
 XX New high-activity endo-xylanase from Streptomyces - used for
 PT treating lignocellulose, pref. pulp for delignification,
 PT brightening and viscosity improvement
 XX
 PS Claim 1; Page 8; 30pp; English.
 XX
 CC Chromosomal DNA was extd. from S. lividans 66 (strain 1326) and

restriction fragments sepd. The strain S. lividans strain 1326 and CC mutated using N-methyl-N'-nitro-N-nitrosoguanidine and a double mutant beta-1,4-D-glucan hydrolase (endocellulase)-negative CC and xylanase-negative was selected. Protocollasing and CC transformation of the double mutant was carried out using the DNA fragments and the multicopy vector pU702 to obtain the xylanase C-harboring multicopy clone xinc. The xylanase enzyme produced by CC this clone has a higher activity than known xylanases and can be used for hydrolysing beta-1,4-D-xylosidic linkages in pulp. The treatment of lignocellulose material with the xylanase results in delignification, brightening and viscosity improvement. Further, such treatment may provide more relaxed fibres resulting in an improved performance or a subsequent treatment, such as swelling, beating, drainage or chemical bleaching of the pulp, with an overall reduction in energy and chemicals used.

(Updated on 25-MAR-2003 to correct PN field.)

Sequence 200 AA;

Query Match 51.9%; Score 41; DB 14; Length 200;
Best Local Similarity 42.9%; Pred. No. 14;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYQNIQY 14
DB 116 SDGCTYDIYQTRRY 129

RESULT 13

AA048535
ID AAB48535 standard; Protein; 240 AA.

AC AAB48535;

DT 05-MAR-2001 (first entry)

DE Streptomyces lividans xylanase C.

KM Bacterial; Bacillus circulans; xylanase; xylanase activity; XA;

KW Bleaching agent.

OS Streptomyces lividans.

PN MO200068396-A2.

PD 16-NOV-2000.

PF 12-MAY-2000; 2000WO-US13172.

PR 12-MAY-1999; 99US-0133714.

PA (XENC-) XENCOR INC.

PI Bentzien JM;

DR WPI; 2000-679800/66.

XX Non naturally occurring XA protein with enhanced thermophilicity;

PT algalophilicity or thermostability relative to the naturally occurring

PT Bacillus circulans xylanase is used in an agent for bleaching pulp -

XX Disclousure; Fig 16D; 114pp; English.

XX The present sequence is given in a specification relating to non
CC naturally occurring xylanase activity (XA) proteins. The XA proteins
CC comprise an amino acid sequence less than 97% identical to a naturally
CC occurring Bacillus circulans xylanase. They are modified to exhibit
CC enhanced thermostability, alkalophilicity or thermostability relative
CC to the naturally occurring B. circulans xylanase. They may be used as
CC the active compound in a bleaching agent which is used for bleaching
CC pulp.

Sequence 240 AA;

Query Match 51.9%; Score 41; DB 21; Length 240;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYQNIQY 14
DB 156 SDGCTYDIYQTRRY 169

RESULT 14

AA018641
ID AA018641 standard; Protein; 240 AA.

AC AA018641;

DT 24-OCT-2002 (first entry)

DE S lividans xylanase C.

KM Xylanase activity; enzyme; thermostable; xylanase; pulp bleaching;

KW liquid clarification; coffee extraction; plant oil extraction;
starch extraction; food thickener; animal food additive; mutant; murein.

OS Streptomyces lividans.

PN MO200238746-A2.

PD 16-MAY-2002.

PF 09-NOV-2001; 2001WO-US48018.

PR 10-NOV-2000; 2000US-0710050.

PA (XENC-) XENCOR INC.

PI Bentzien J; Dahlyat B;

DR WPI; 2002-608200/65.

XX Novel xylanase activity protein, useful in bleaching process of pulp

PT and in food and animal feed industry, has enhanced thermostability and

PT alkalophilicity -

PS Disclousure; Fig 16D; 121pp; English.

XX The present invention relates to a non-naturally occurring xylanase
CC activity (XA) protein comprising an amino acid sequence less than 97%
CC identical to a naturally occurring Bacillus circulans xylanase, where the
CC protein has been modified to exhibit enhanced thermostability,
CC alkalophilicity, or thermostability relative to naturally occurring B.
CC circulans xylanase, and has at least 5 amino acid substitutions. A
CC bleaching agent comprising a modified xylanase is useful for bleaching
CC pulp, in the bioconversion of lignocellulosic materials to fuels, for
CC clarifying juice and wine, extracting coffee, plant oils and starch,
CC producing food thickeners, altering texture in bakery products, e.g.
CC improving the quality of dough, helping bread to rise and processing of
CC wheat and corn for starch production, use as animal food additives to aid
CC in the digestibility of feedstuffs and in the washing of super precision
CC devices and semiconductors. The present sequence is a xylanase protein
CC described in the exemplification of the invention.

XX Sequence 240 AA;

Query Match 51.9%; Score 41; DB 23; Length 240;
Best Local Similarity 42.9%; Pred. No. 18;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 AEGSTXDYVYQNIQY 14
DB 156 SDGCTYDIYQTRRY 169

RESULT 15

ABG13264
ID ABG13264 standard; Protein: 359 AA.

XX ABG13264;

AC 18-FEB-2002 (first entry)

DT Novel human diagnostic protein #13255.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM Food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

PD 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

PR (HYSE-) HYSEQ INC.

XX Dmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

XX N-PSDB; AAS77451.

DR New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics; forensics; gene mapping; identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX Claim 20; SEQ ID No 43623; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG0377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 359 AA;

Query Match 51.9%; Score 41; DB 22; Length 359;

Best Local Similarity 46.7%; Pred. No. 30; Mismatches 0; Gaps 0;

Matches 7; Conservative 3; Indels 0;

QY 2 EGSTXDYONIOYAG 16

DB 46 DGFATFIQIKITLG 60

Search completed: January 29, 2004, 14:44:28
Job time: 42 secs